Bagging the Network*

Ming Li Zhentao Shi Yapeng Zheng†

April 27, 2026

Abstract

We develop a unified estimation and inference framework for dyadic network formation with individual fixed effects, covering both transferable-utility (TU) and nontransferable-utility (NTU) links under general link functions. Under NTU, bilateral consent makes the fixed effects non-additive and the log-likelihood non-concave in the high-dimensional fixed effects, so differencing and profile-likelihood methods fail. We combine a joint method-of-moments initial estimator, a Le Cam one-step refinement, and a split-network jackknife bagging step that removes the incidental parameter bias without inflating variance. The resulting homophily estimator is asymptotically normal, unbiased, and attains the Cramér–Rao lower bound without requiring the log-likelihood to be concave in the fixed effects; we extend the theory to average partial effects and establish robustness to link-function misspecification. Simulations under both TU and NTU designs confirm these predictions. Applied to Thai village networks (TU), kinship and wealth differences both increase linking; in the Nyakatoke risk-sharing network (NTU), wealth differences have no significant effect, mirroring the two regimes' distinct logics.

Keywords: bootstrap aggregating, dyadic network formation, transferable utilities, nontransferable utilities, one-step approximation, fixed effects.

*We thank Donald Andrews, Xiaohong Chen, Chih-Sheng Hsieh, Bryan Graham, Matthew O. Jackson, Yuichi Kitamura, Oliver Linton, Shuyang Sheng, Martin Weidner, Weichen Wang, Jun Yu, Yichong Zhang, and seminar participants at the University of Sydney, University of Macau, Singapore Management University, Singapore Workshop on Econometrics and Data Sciences, Asia Meeting of the Econometric Society (NYU Abu Dhabi), Cowles Conference on Econometrics (Yale), and Econometric Society World Congress for comments.

†Li: Department of Economics and Risk Management Institute, National University of Singapore, `mli@nus.edu.sg`. Shi (corresponding author): Department of Economics, The Chinese University of Hong Kong, `zhentao.shi@cuhk.edu.hk`. Zheng: Department of Economics, The Chinese University of Hong Kong, `yapengzheng@link.cuhk.edu.hk`.

1 Introduction

Dyadic network formation (modeling how links arise between pairs of agents) is central to understanding social and economic phenomena such as friendship networks, risk-sharing arrangements, and inter-firm alliances. In such models, individual unobserved heterogeneity is pervasive: agents differ in sociability, ability, or other latent traits that simultaneously affect their desirability as partners and correlate with observed covariates. Treating this heterogeneity as individual fixed effects is therefore essential, yet it poses econometric challenges, most notably the incidental parameter problem, which leads to asymptotic bias and substantial computational burdens when the number of fixed effects grows with the network size. Whether fixed-effects inference for dyadic network-formation models can be pushed to Cramér–Rao efficiency under *general* link functions, particularly for bilateral-consent settings where fixed effects enter non-additively, has remained an open question, which this paper closes.

Whether a link involves transferable or nontransferable utilities has direct consequences for econometric analysis. Under transferable utilities (TU), the total surplus from a link is additively separable in the two agents' fixed effects, and a range of methods exploit this separability and specific distributional assumptions on the idiosyncratic shocks to eliminate or profile out the fixed effects (Chatterjee, Diaconis, and Sly, 2011; Graham, 2017; Dzemski, 2019). A method that accommodates general link functions is still lacking in the literature. Furthermore, many economically important networks involve nontransferable utilities (NTU): friendships form when both individuals are willing, risk-sharing arrangements require mutual consent, and bilateral trade links are sustained only when each party individually benefits. Under NTU, each agent's latent type enters the linking probability through a separate agreement condition, so the fixed effects are no longer additively separable and the Bernoulli log-likelihood is no longer concave in them. This non-additivity and non-concavity simultaneously break the differencing and profile-likelihood methods developed for TU, all of which rely on either additive bias structure or global concavity of the objective in the nuisance parameters. Despite the empirical prevalence of NTU settings, no existing framework delivers inference for homophily parameters or consistent estimation of fixed effects under NTU with a general link function. Bilateral consent without side payments characterizes a significant proportion of networks studied in applied research, including friendships, risk-sharing, collaboration, and information exchange. This gap is therefore consequential: applied researchers currently have no available procedure to compute standard errors and conduct hypothesis tests for precisely the class of networks they routinely encounter.

This paper proposes a unified parametric framework for estimation and inference under both TU and NTU with general link functions, based on a single large network with observed pairwise covariates and individual fixed effects.[1] Our estimation strategy proceeds in three steps. First, we construct a joint method-of-moments (JMM) estimator for the homophily parameters β_0 and the fixed effects $\boldsymbol{\alpha}_0$, and establish consistency for both and normality for the homophily parameter. Second, we refine the JMM estimator of β_0 via Le Cam (1969)'s one-step approximation to the maximum likelihood estimator (MLE), which requires merely a single Newton-type update and avoids the computational burden and numerical instability of full maximum likelihood optimization with high-dimensional fixed effects. Third, we debias the one-step estimator, which inherits the asymptotic bias of the MLE, through a split-network jackknife combined with bootstrap aggregating (bagging) (Breiman, 1996; Hirano and Wright, 2017), a technique that is novel in the network formation literature.

Our contributions are threefold. *Theoretically*, we deliver the first asymptotic inference and efficiency theory for homophily parameters under NTU, with ℓ_∞-consistent estimation of the individual fixed effects; the bagging estimator $\widehat{\beta}_{\mathrm{BG}}$ is asymptotically normal, asymptotically unbiased, attains the Cramér–Rao lower bound (CRLB, Rao, 1992), and is robust to the choice of random splits. *Methodologically*, our approach sidesteps log-likelihood concavity altogether: JMM solves a degree-matching moment condition rather than maximizing the likelihood, while the one-step and bagging refinements use the outer-product Fisher information, which is positive definite by construction. Concavity fails structurally under NTU and under TU with non-log-concave shock densities; our framework is the first to deliver Cramér–Rao-efficient inference in both cases. We also establish asymptotic normality for the average partial effects (APEs), robustness to link-function misspecification, and a multi-network extension. *Empirically*, we provide the first side-by-side evidence of how the TU vs. NTU distinction shapes estimated homophily: wealth differences drive linking under TU (Thai villages) but not under NTU (Nyakatoke).

Efficiency and asymptotic unbiasedness matter here because they determine whether applied work can distinguish economically meaningful effects from noise.

[1] We provide a more in-depth comparison between TU and NTU in Remark 1.

An inefficient estimator yields unnecessarily wide confidence intervals; a biased estimator—even if $\sqrt{N}$-consistent—invalidates plug-in hypothesis tests. Our bagging estimator is simultaneously unbiased and CRLB-efficient, yielding valid confidence intervals as narrow as the model allows.

Simulation results confirm that the proposed estimators for the homophily parameters, individual fixed effects, and APEs perform as predicted by the theory under both the NTU baseline and a TU counterpart using either logistic or probit link functions. In particular, the non-concavity of the log-likelihood renders MLE unreliable for estimating the high-dimensional fixed effects $\boldsymbol{\alpha}_0$: MLE produces substantially larger root mean squared errors (RMSEs) than JMM, with the failure most pronounced under bimodal or hub–periphery designs (Figure C.1 in the Supplemental Material). We present two empirical examples that exhibit the TU and NTU regimes. First, under TU, we apply our method to the Townsend Thai village networks (Kinnan, Samphantharak, Townsend, and Vera-Cossio, 2024), pooling across 16 villages to estimate link formation for financial, operations, and labor transactions. We find that kinship is a strong and broadly positive predictor across all three networks, and that greater absolute net-worth differences between households are also associated with more frequent linking, consistent with gains-from-trade in transactional relationships under TU. Second, under NTU, we revisit the Nyakatoke risk-sharing network (De Weerdt, 2004), where our method indicates that wealth differences have no statistically significant effect on link formation. The contrast between these two settings is not incidental: TU surplus-sharing predicts that wealth heterogeneity *fuels* linking in financial, operations, and labor transactions (as in the Thai estimates), while the bilateral-consent logic of risk pooling predicts that wealth heterogeneity does *not* drive linking when both sides must consent (as in the Nyakatoke estimate). The same econometric framework recovers both patterns without imposing the separability that distinguishes them. A code-and-data demonstration of our proposed methods is available at the GitHub repository `https://github.com/YapengZheng/bagging_network`.

Literature Review. Our paper contributes to the literature on dyadic network formation in a single large network. Most existing work studies TU, which allows individual fixed effects to be eliminated by arithmetic differencing (Graham, 2017; Zeleneev, 2026; see Graham, 2020, for a review). Gao, Li, and Xu (2023) study a semiparametric model under NTU using logical differencing, but without inference for homophily parameters or estimators of fixed effects. We complement their work by establishing inference for homophily parameters, delivering ℓ_∞-consistent estimators of fixed effects, and developing asymptotic results for the APEs.

Our paper also builds on Graham (2017), who introduces a tetrad logit estimator and a joint MLE under TU and logistic link functions. These methods, as well as functional differencing (Bonhomme, 2012), do not extend to NTU because the fixed effects enter the linking probability non-additively (as $p(\alpha_i, \alpha_j, x_{ij}^\top\beta)$ rather than $p(\alpha_i + \alpha_j + x_{ij}^\top\beta)$). Recent contributions also remain within TU: Hughes (2026) develops a jackknife bias correction, Qu, Chen, Yan, and Chen (2025) propose a projection approach for directed networks, and Gao (2020) study semiparametric estimation. In contrast, our estimator applies under both TU and NTU. An earlier working paper by Shi and Chen (2016) analyzed the MLE for this model under NTU. The present paper subsumes that analysis by establishing formal efficiency via the Le Cam one-step approximation, developing a stable procedure that avoids the non-concavity of direct MLE (see Section 2 and Figure C.1 in the Supplemental Material), and introducing bagging for debiasing.

Methodologically, our work relates to the large-T panel literature on nonlinear fixed-effect models (Hahn and Newey, 2004; Dhaene and Jochmans, 2015; Fernández-Val and Weidner, 2016, 2018; Honoré and de Paula, 2021). These methods rely on concavity of log-likelihood functions and/or sparsity assumptions on certain derivatives of functionals of the fixed effects that are hard to verify in our setting. Instead, we adapt the sample-splitting idea (see Mei, Sheng, and Shi, 2026; Liao, Mei, and Shi, 2024 for tackling Nickell-type biases in panel predictive regressions using split-sample strategies) and establish formally that bagging delivers unbiased and efficient estimation. Related work on orthogonalized estimators (Bonhomme, Jochmans, and Weidner, 2024) requires an additive bias structure that fails under NTU. The Bernoulli log-likelihood is non-additive because the $\log(1 - p_{ij})$ term involves the product $F(\alpha_i + x_{ij}^\top\beta)\, F(\alpha_j + x_{ij}^\top\beta)$, which violates the additive bias structure required by their Assumption 3.

Finally, there is a line of work on strategic network formation and empirical games based on pairwise stability (e.g., Jackson and Wolinsky, 1996; de Paula, Richards-Shubik, and Tamer, 2018). These models incorporate externalities but typically impose restrictions on heterogeneity or the degree distribution and often require TU; Gao, Li, and Xu (2026) recently provide tractable identification in strategic models with unobserved heterogeneity under TU. Our fixed-effects approach, which accommodates both TU and NTU and permits arbitrary correlation between observables and the fixed effects, is therefore complementary to and methodologically distinct from the existing literature (for a review of the two approaches, see de Paula, 2020a).

Organization. Section 2 states the model and estimation algorithm; Section 3 develops the asymptotic theory; Section 4 covers APEs and link-function misspecification; Sections 5 and 6 report simulations and empirical applications; Section 7 concludes. Appendices collect notation, supporting lemmas, and proofs; the Supplemental Material contains proofs of supporting lemmas, the multi-network extension, extended simulations, and data descriptions.

Notation. Let ":=" denote a definition, and let the superscript "$\top$" denote the transpose of a vector or a matrix. We use boldface for variables of increasing dimension with n. For example, the true fixed effects $\boldsymbol{\alpha}_0 = (\alpha_{i0})_{1\le i\le n}$ are $n \times 1$. For an $n \times 1$ vector $\mathbf{a} = (a_1, \ldots, a_n)^\top$, its ℓ_1 norm is $\|\mathbf{a}\|_1 := \sum_{i=1}^n |a_i|$, ℓ_2 norm is $\|\mathbf{a}\|_2 := (\sum_{i=1}^n a_i^2)^{1/2}$, and ℓ_∞ norm is $\|\mathbf{a}\|_\infty := \max_{1\le i\le n} |a_i|$. When $O(\cdot)$ (and other notation for order) is written for a vector (or matrix), it means that each element in the vector (or matrix) is of the order in $O(\cdot)$. Here, "plim" denotes the probability limit, "$\xrightarrow{p}$" convergence in probability, and "$\xrightarrow{d}$" convergence in distribution. Unless otherwise noted, for all convergence results we pass $n \to \infty$. For an $n \times n$ matrix $\mathbf{A}$, we write $\|\mathbf{A}\|_1 := \max_{1\le i\le n} \|\mathbf{A}_{\cdot i}\|_1$, $\|\mathbf{A}\|_\infty := \max_{1\le i\le n} \|\mathbf{A}_{i\cdot}\|_1$ and $\|\mathbf{A}\|_{\max} := \max_{1\le i,j\le n} |\mathbf{A}_{ij}|$, where $\mathbf{A}_{\cdot i}$ and $\mathbf{A}_{i\cdot}$ are the ith column and row of $\mathbf{A}$, respectively. To simplify notation, we write $p_{ij}(\boldsymbol{\alpha}, \beta) := p(\alpha_i, \alpha_j, x_{ij}^\top \beta)$. The bare form p_{ij} denotes this function with the arguments $(\boldsymbol{\alpha}, \beta)$ suppressed; we write $p_{ij,0} := p_{ij}(\boldsymbol{\alpha}_0, \beta_0)$ for its value at the true parameters. Evaluation at any other specific arguments is written out explicitly. The same convention applies to other objects that are functions of $(\boldsymbol{\alpha}, \beta)$. Finally, the abbreviation "w.p.a.1" stands for "with probability approaching one."

2 Model and Computation

We consider an undirected network formed among agents $i \in \mathcal{I}_n := \{1, \ldots, n\}$. Hence, there are $N := \binom{n}{2}$ dyads to be linked. An observed link Y_{ij} between i and j

is formed with probability:

$$\Pr(Y_{ij}=1|X_{ij},\alpha_{i0},\alpha_{j0}) \coloneqq p(\alpha_{i0},\alpha_{j0},X_{ij}^\top\beta_0) \text{ for } 1\le i\ne j\le n, \tag{1}$$

where $p(\cdot):\mathbb{R}^3\to(0,1)$ is a user-specified symmetric function in α_i and α_j. We rule out self-loops, i.e., $Y_{ii}=0,\ i\in\mathcal{I}_n$. The link probability $p_{ij,0}$ is jointly determined by the two agent-specific fixed effects $(\alpha_{i0},\alpha_{j0})$ and the dyad-specific index $X_{ij}^\top\beta_0$ that captures the homophily effect in the observable characteristics of each pair (i,j), where $X_{ij}\in\mathbb{R}^K$ denotes the symmetric dyad-level covariates for all $i\ne j$. The log-likelihood is

$$\ell_n(\boldsymbol{\alpha},\beta) \coloneqq \sum_{i=1}^n\sum_{j>i}\{Y_{ij}\log p_{ij}(\boldsymbol{\alpha},\beta)+(1-Y_{ij})\log(1-p_{ij}(\boldsymbol{\alpha},\beta))\}.$$

Remark 1. Our network formation model (1) covers both TU and NTU:

$$Y_{ij}=\mathbb{1}\left\{\alpha_{i0}+\alpha_{j0}+X_{ij}^\top\beta_0-\epsilon_{ij}>0\right\} \text{ and} \tag{TU}$$

$$Y_{ij}=\mathbb{1}\left\{\alpha_{i0}+X_{ij}^\top\beta_0-\epsilon_{ij}>0\right\}\mathbb{1}\left\{\alpha_{j0}+X_{ji}^\top\beta_0-\epsilon_{ji}>0\right\}, \tag{NTU}$$

where ϵ_{ij} is an idiosyncratic error with a known distribution. The model (TU) essentially asserts that, if the joint surplus generated by a bilateral link $\alpha_{i0}+\alpha_{j0}+X_{ij}^\top\beta_0-\epsilon_{ij}$ is positive, then the link between i and j is formed. An important assumption behind the model (TU) is that the link surplus can be freely distributed between i and j, and that bargaining efficiency is always achieved, which is a strong assumption in many networks (e.g., risk-sharing networks and friendship networks). The model (NTU), on the other hand, requires that the utility surplus from the link for both i and j be strictly positive in order to form a link, which is arguably more realistic in the aforementioned networks. Furthermore, the model (NTU) reflects the fact that the party with relatively lower utility is the pivotal one in link formation. Finally, it can be shown (see Gao, Li, and Xu, 2023) that the model (NTU) can accommodate homophily effects in both observable and unobservable covariates.

Given the model, we introduce the algorithm to estimate the homophily coefficient β_0. There are three sequential modules—JMM, one-step (OS), and Bagging (BG)—that lead to $\widehat{\beta}_{\mathrm{BG}}$. Specifically, Module JMM provides an initial consistent estimator, which, however, does not reach the CRLB and is biased. We refine the JMM estimator with the one-step adjustment to achieve the CRLB. Finally, we apply the bagged split-network jackknife to debias the one-step estimator while preserving its efficiency.

We define a few objects before each module. Let $\mathbf{Y} = (Y_{ij})_{1\le i,j\le n}$ be the $n \times n$ adjacency matrix and $\mathbf{X} = (X_{ij})_{1\le i,j\le n}$ be the $n \times n \times K$ random tensor of covariates. Denote their realizations by $\mathbf{y} = (y_{ij})_{1\le i,j\le n}$ and $\mathbf{x} = (x_{ij})_{1\le i,j\le n}$, respectively. The *degree* $d_i \coloneqq \sum_{j\ne i} y_{ij}$ is defined for each $i \in \mathcal{I}_n$ of the observed network $\mathbf{Y}$. Define a vector of moment functions $\mathbf{m}(\boldsymbol{\alpha}, \beta) \coloneqq (\mathbf{m}_1^\top(\boldsymbol{\alpha}, \beta), m_2^\top(\boldsymbol{\alpha}, \beta))^\top$, where $\mathbf{m}_1(\boldsymbol{\alpha}, \beta) \coloneqq (d_i - \sum_{j\ne i} p_{ij}(\boldsymbol{\alpha}, \beta))_{i=1}^n$ is an n-dimensional function that concerns the average degree of each i, and $m_2(\boldsymbol{\alpha}, \beta) \coloneqq \sum_{i=1}^{n} \sum_{j>i} [y_{ij} - p_{ij}(\boldsymbol{\alpha}, \beta)] x_{ij}$ is a K-dimensional function.

Module 1 (JMM)**.** The **JMM estimator** $(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})$ is the solution to the $(n + K)$-equation system $\mathbf{m}(\boldsymbol{\alpha}, \beta) = 0$.

The JMM estimator is just-identified with $n + K$ moment conditions for $n + K$ unknowns, and we choose degree-based equations in $\mathbf{m}_1$ for their simplicity and stability (Graham, 2017). To find the solution to $\mathbf{m}(\boldsymbol{\alpha}, \beta) = 0$, for each β we let

$$r_i(\boldsymbol{\alpha}, \beta) = \alpha_i + (n-1)^{-1}\left(d_i - \sum_{j\ne i} p_{ij}(\boldsymbol{\alpha}, \beta)\right), \qquad i \in \mathcal{I}_n \tag{2}$$

and $\mathbf{r}(\boldsymbol{\alpha}, \beta) = \big(r_1(\boldsymbol{\alpha}, \beta), \ldots, r_n(\boldsymbol{\alpha}, \beta)\big)^\top$. The intuition is, for any i when d_i is strictly larger than $\sum_{j\ne i} p_{ij}(\boldsymbol{\alpha}, \beta)$, we would like to increase α_i such that each $p_{ij}(\boldsymbol{\alpha}, \beta)$ for $j \ne i$ is larger, and vice versa. Starting with an initial value $\boldsymbol{\alpha}^0$, we iterate $\boldsymbol{\alpha}^{k+1}(\beta) = \mathbf{r}(\boldsymbol{\alpha}^k(\beta), \beta)$ until convergence to obtain $\widehat{\boldsymbol{\alpha}}(\beta)$, and then we solve the finite dimensional equations $m_2(\widehat{\boldsymbol{\alpha}}(\beta), \beta) = 0$. Unlike existing methods (e.g., Theorem 1.5 of Chatterjee, Diaconis, and Sly, 2011; equation (17) of Graham, 2017), we do not rely on specific functional-form assumption on the idiosyncratic shocks (e.g., logit or probit): $\widehat{\boldsymbol{\alpha}}(\beta)$ is recovered purely through a contraction-type fixed-point argument on (2).

The OS module involves the score and information matrix. Define the score of ℓ_n as $\mathbf{s}(\boldsymbol{\alpha}, \beta) = \big(\mathbf{s}_1^\top(\boldsymbol{\alpha}, \beta), s_2^\top(\boldsymbol{\alpha}, \beta)\big)^\top = \big(\partial \ell_n/\partial \boldsymbol{\alpha}^\top, \partial \ell_n/\partial \beta^\top\big)^\top$, and partition the information matrix

$$\mathbf{I}(\boldsymbol{\alpha}, \beta) = \mathbb{E}[\mathbf{s}(\boldsymbol{\alpha}, \beta)\mathbf{s}(\boldsymbol{\alpha}, \beta)^\top | \mathbf{x}, \boldsymbol{\alpha}] =: \begin{pmatrix} \mathbf{I}_{11}(\boldsymbol{\alpha}, \beta) & \mathbf{I}_{12}(\boldsymbol{\alpha}, \beta) \\ \mathbf{I}_{12}^\top(\boldsymbol{\alpha}, \beta) & \mathrm{I}_{22}(\boldsymbol{\alpha}, \beta) \end{pmatrix} \tag{3}$$

into four compatible blocks. Define the concentrated score function and information matrix of β as

$$s_n(\boldsymbol{\alpha}, \beta) = s_2(\boldsymbol{\alpha}, \beta) - \mathbf{I}_{12}(\boldsymbol{\alpha}, \beta)^\top \mathbf{I}_{11}(\boldsymbol{\alpha}, \beta)^{-1} \mathbf{s}_1(\boldsymbol{\alpha}, \beta) \text{ and}$$

$$\mathrm{I}_n(\boldsymbol{\alpha},\beta) = \mathrm{I}_{22}(\boldsymbol{\alpha},\beta) - \mathbf{I}_{12}(\boldsymbol{\alpha},\beta)^\top \mathbf{I}_{11}(\boldsymbol{\alpha},\beta)^{-1}\mathbf{I}_{12}(\boldsymbol{\alpha},\beta), \tag{4}$$

respectively. Here, $s_n(\boldsymbol{\alpha},\beta)$ and $\mathrm{I}_n(\boldsymbol{\alpha},\beta)$ are the concentrated counterparts of $s_2(\boldsymbol{\alpha},\beta)$ and $\mathrm{I}_{22}(\boldsymbol{\alpha},\beta)$ after concentrating out the influence from estimating $\boldsymbol{\alpha}$, respectively.

Module 2 (OS)**.** Substitute the JMM estimator $(\widehat{\boldsymbol{\alpha}},\widehat{\beta})$ into

$$\widehat{\beta}_{\mathrm{OS}} \coloneqq \widehat{\beta} + \mathrm{I}_n(\widehat{\boldsymbol{\alpha}},\widehat{\beta})^{-1} s_n(\widehat{\boldsymbol{\alpha}},\widehat{\beta}). \tag{5}$$

Module OS is Le Cam (1969)'s one-step approximation of the MLE: starting from the $\sqrt{N}$-consistent JMM estimate, a single Newton update along the efficient-score direction closes the efficiency gap and delivers the CRLB asymptotically. In (5), s_n is the slope of the log-likelihood in β after profiling out $\boldsymbol{\alpha}$, and I_n^{-1} rescales the step by the inverse curvature, so the adjustment is larger when the likelihood is flat in β and smaller when it is sharply curved. Module OS sidesteps the non-concavity and numerical instability of direct MLE optimization (see Remark 4).

Finally, the bagging step introduces randomization. Assume an even integer n for convenience. Let $t = 1, 2, \ldots, \widetilde{T}_n$, for some $\widetilde{T}_n \leq \binom{n}{n/2}$, index an equal-sized random partition of $\mathcal{I}_n$ into $\mathcal{I}_{1,n}^{(t)}$ and $\mathcal{I}_{2,n}^{(t)}$ such that $\mathcal{I}_{1,n}^{(t)} \cup \mathcal{I}_{2,n}^{(t)} = \mathcal{I}_n$, $\mathcal{I}_{1,n}^{(t)} \cap \mathcal{I}_{2,n}^{(t)} = \emptyset$, and the splits are independent over t.

Module 3 (BG)**.** For each $t = 1, \ldots, \widetilde{T}_n$, hold the full-sample JMM estimator $\widehat{\beta}$ fixed and, on the subnetwork indexed by $\mathcal{I}_{1,n}^{(t)}$, update only $\widehat{\boldsymbol{\alpha}}$ via the fixed-point iteration in (2). Apply the OS step on that split to obtain $\widehat{\beta}_{\mathrm{OS},1}^{(t)}$ using the new $\widehat{\boldsymbol{\alpha}}$ and full-sample $\widehat{\beta}$. Repeat the same procedure on $\mathcal{I}_{2,n}^{(t)}$ to obtain $\widehat{\beta}_{\mathrm{OS},2}^{(t)}$. Apply the bagged jackknife to obtain the **BG estimator**

$$\widehat{\beta}_{\mathrm{BG}} \coloneqq 2\widehat{\beta}_{\mathrm{OS}} - (2\widetilde{T}_n)^{-1}\sum_{t=1}^{\widetilde{T}_n}(\widehat{\beta}_{\mathrm{OS},1}^{(t)} + \widehat{\beta}_{\mathrm{OS},2}^{(t)}).$$

We employ split-network jackknife to debias $\widehat{\beta}_{\mathrm{OS}}$. Due to the equal splits of nodes, each of $\widehat{\beta}_{\mathrm{OS},1}^{(t)}$ and $\widehat{\beta}_{\mathrm{OS},2}^{(t)}$ incurs twice the leading bias in the asymptotic expansion. If we apply the split-network jackknife only once, then

$$\widehat{\beta}_{\mathrm{OS-SJ}}^{(t)} \coloneqq 2\widehat{\beta}_{\mathrm{OS}} - \frac{1}{2}\left(\widehat{\beta}_{\mathrm{OS},1}^{(t)} + \widehat{\beta}_{\mathrm{OS},2}^{(t)}\right)$$

self-cancels the leading bias. However, the variance of $\widehat{\beta}_{\mathrm{OS-SJ}}^{(t)}$ is doubled, since splitting the network in half causes the links between nodes belonging to different subnetworks to be ignored. Furthermore, splitting the whole network randomly makes the estimator computationally unstable. To deal with these issues, we let $\widetilde{T}_n \to \infty$ and indeed $\widehat{\beta}_{\mathrm{BG}} = \widetilde{T}_n^{-1} \sum_{t=1}^{\widetilde{T}_n} \widehat{\beta}^{(t)}_{\mathrm{OS-SJ}}$ averages $\widehat{\beta}^{(t)}_{\mathrm{OS-SJ}}$ over $\widetilde{T}_n$ independent splits.

Remark 2. It is asymptotically valid if we split the nodes into $G \geq 2$ equal parts and form a jackknife-type correction with appropriate weights, as discussed in Dhaene and Jochmans (2015) for panel data. However, increasing G reduces the number of nodes in each subsample, which degrades the quality of fixed-effect estimates and can cause numerical instability. Our approach instead fixes $G = 2$, the minimal split needed for first-order bias correction, and aggregates over $\widetilde{T}_n \to \infty$ independent random splits. This bagging strategy is crucial: a single split-network jackknife eliminates the leading bias but doubles the asymptotic variance, whereas averaging over many splits deflates the variance back to the Cramér–Rao lower bound (Theorem 3).

When computing $\widehat{\beta}^{(t)}_{\mathrm{OS},1}$ and $\widehat{\beta}^{(t)}_{\mathrm{OS},2}$ for each random split t, we do not recompute a split-specific initial JMM estimate of β. Instead, the full-sample JMM estimator $\widehat{\beta}$ is retained, $\widehat{\boldsymbol{\alpha}}$ is re-estimated on each split via (2), and the OS update then produces the split-specific estimators $\widehat{\beta}^{(t)}_{\mathrm{OS},1}$ and $\widehat{\beta}^{(t)}_{\mathrm{OS},2}$. The rationale is that the incidental parameter bias of the OS estimator comes from the estimation of high-dimensional fixed effects: $\widehat{\alpha}_i$ converges at rate $\sqrt{(\log n)/n}$, which is slower than $\sqrt{N} \asymp n$ for $\widehat{\beta}$, and it is this slower rate that generates the $O(N^{-1/2})$ bias requiring correction. Consequently, the procedure remains computationally efficient for moderate values of $\widetilde{T}_n$, such as 200 or 400 in our simulations.

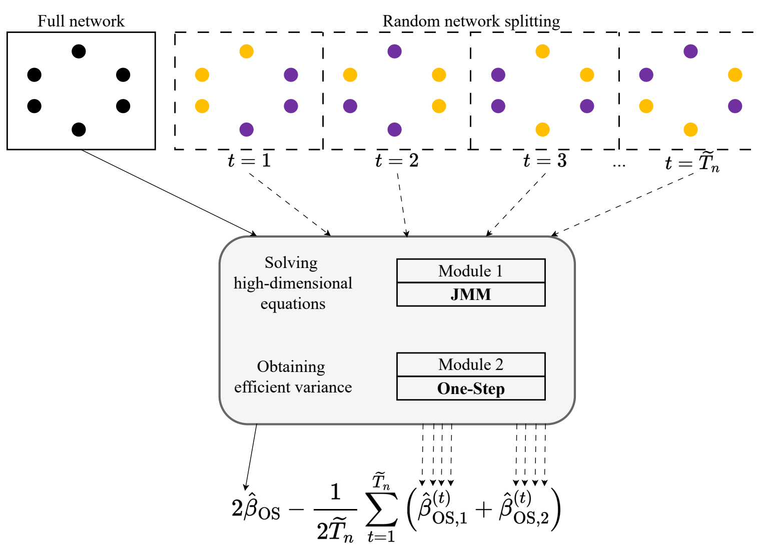

Figure 1: Flowchart of the estimation procedure

Figure 1 summarizes the procedure. Modules JMM and OS are first applied to the full network (black dots) to obtain $\widehat{\beta}_{\mathrm{OS}}$. For each random split $t = 1, \ldots, \widetilde{T}_n$, the nodes are divided into two halves (e.g., yellow and purple dots in split $t = 1$). On each half, JMM re-estimates $\boldsymbol{\alpha}$ using the full-sample $\widehat{\beta}$, and the OS step updates $\widehat{\beta}$ to yield $\widehat{\beta}^{(t)}_{\mathrm{OS},1}$ and $\widehat{\beta}^{(t)}_{\mathrm{OS},2}$. Averaging the bagged jackknife over all splits produces $\widehat{\beta}_{\mathrm{BG}}$.

3 Large Sample Theory

This section proves that the bagging estimator $\widehat{\beta}_{\mathrm{BG}}$ is asymptotically normal, unbiased, and CRLB-efficient. We reach this result in three steps. Section 3.1 shows that the JMM estimator is $\sqrt{N}$-consistent for β_0 and that the plug-in $\widehat{\boldsymbol{\alpha}}(\widehat{\beta})$ converges to $\boldsymbol{\alpha}_0$ in the ℓ_∞ norm, enough to enable the one-step correction. Section 3.2 shows that the one-step estimator $\widehat{\beta}_{\mathrm{OS}}$ attains the CRLB but carries an $O(n^{-1})$ incidental-parameter bias. Section 3.3 then shows that bagging removes this bias without inflating the asymptotic variance.

We first state four baseline assumptions that underlie the theoretical results.

Assumption 1 (Correctly Specified Model)**.** *The conditional likelihood of* $\mathbf{Y} = \mathbf{y}$ *given* $\mathbf{X} = \mathbf{x}$ *and* $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ *is*

$$\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}, \boldsymbol{\alpha} = \boldsymbol{\alpha}_0) = \prod_{i=1}^{n} \prod_{j>i} p_{ij,0}^{y_{ij}} \left(1 - p_{ij,0}\right)^{1-y_{ij}}.$$

Assumption 1 is similar to Assumption 1 of Graham (2017), except for two important differences. First, under potentially NTU α_{i0} and α_{j0} are not additively separable in the linking probability between i and j, and thus the tetrad logit estimator of Graham (2017) does not apply in our setting. Second, the functional form of $p(\cdot)$ is general (subject to Assumption 4 below) and includes the commonly used logistic (e.g., Chatterjee, Diaconis, and Sly, 2011; Graham, 2017; Qu, Chen, Yan, and Chen, 2025) and probit as special cases. The multiplicative form of the joint likelihood implies that the links are formed independently of one another conditional on the agent attributes. It is suitable in settings such as risk-sharing networks, online friendships, trade networks, and conflicts between nation-states. Conversely, settings with strategic link formation, where agents' linking decisions depend on other links in the network, fall outside this framework.

Assumption 1 also requires the link function $p(\cdot)$ to be correctly specified. It is well-known that, under regularity conditions, the MLE converges to the parameter value that minimizes the Kullback-Leibler divergence between the true and the misspecified model. The issue is complicated by the high-dimensional individual fixed effects and has not been investigated in the network formation literature. In the Supplemental Material, we discuss the impact of link function misspecification on the theoretical results in Section B and provide supporting simulation evidence in Section C.

Assumption 2 (Bounded Support)**.**

(i) $\boldsymbol{\alpha}_0$ *lies in the interior of a compact set* $\mathbb{A} \subset \mathbb{R}^n$.

(ii) β_0 *lies in the interior of a compact set* $\mathbb{B} \subset \mathbb{R}^K$.

(iii) X_{ij} *satisfy* $X_{ij} \in \mathbb{X} \subset \mathbb{R}^K$ *for some compact set* $\mathbb{X}$.

Assumption 2 collects and combines Graham (2017)'s Assumptions 2 and 5(i). Together with Assumption 4 below, it implies that the probability of a link forming between dyad (i, j) is uniformly bounded within $[\kappa, 1-\kappa]$ for some $\kappa \in (0, 1/2)$, which requires the network to be dense.[2] The dense network makes it possible to estimate α_{i0} consistently for each i.

Note that our theory in principle can allow the support of X_{ij} to be unbounded; however, it would add little theoretical insight but incur more technical complexity in the rates of convergence via the Bernstein inequalities to bound the tail probabilities of random variables. Assumption 2(iii), which is similar to Graham (2017, Assumption 2(ii)), allows us to focus on the main idea.

Assumption 3 (Random Sampling)**.** *The sequence* $\{(\alpha_{i0}, X_i)\}_{i=1}^n$ *is an i.i.d. sample from a Borel probability measure* ν *on* $\mathbb{A} \times \mathbb{X}_0$*, where* $\mathbb{X}_0$ *is a compact subset of a finite-dimensional Euclidean space. The dyadic covariate satisfies* $X_{ij} = h(X_i, X_j)$ *for a fixed measurable symmetric function* $h : \mathbb{X}_0 \times \mathbb{X}_0 \to \mathbb{X}$*, where* $\mathbb{X}$ *is the compact set in Assumption 2(iii).*

Assumption 3 parallels Graham (2017, Assumption 3): an agent is an i.i.d. draw from a population with unrestricted joint distribution over (α_{i0}, X_i), accommodating

[2] Density of an undirected network is defined as $\rho_n = N^{-1} \sum_{i=1}^n \sum_{j>i} y_{ij}$. A network is dense if $\lim_{n\to\infty} \rho_n \in [c_1, c_2]$ for some constant $0 < c_1 \leq c_2 < 1$.

arbitrary correlation between unobserved heterogeneity and observed covariates. This is a "fixed-effects" treatment in the sense of Chamberlain (1984), and provides the regularity needed for the limiting objects in Lemma 5 of Appendix B to be well-defined.

For $r_1, r_2, r_3 \in \mathbb{N} \cup \{0\}$, define

$$p^{(r_1,r_2,r_3)}(\alpha_i, \alpha_j, t) := \frac{\partial^{r_1+r_2+r_3} p(\alpha_i, \alpha_j, t)}{\partial \alpha_i^{r_1} \partial \alpha_j^{r_2} \partial t^{r_3}}.$$

Analogously to $p_{ij}(\boldsymbol{\alpha}, \beta)$, we write $p_{ij}^{(r_1,r_2,r_3)}(\boldsymbol{\alpha}, \beta) := p^{(r_1,r_2,r_3)}(\alpha_i, \alpha_j, x_{ij}^\top \beta)$.

Assumption 4 (Restrictions on $p(\cdot)$)**.** *The link-probability function $p(\alpha_i, \alpha_j, t)$ is symmetric in (α_i, α_j) and three times continuously differentiable in (α_i, α_j, t). There exist constants $c_1 \in (0, 1/2]$ and $c_2, c_3 > 0$ such that, for all $(\boldsymbol{\alpha}, \beta) \in \mathbb{A} \times \mathbb{B}$, $x_{ij} \in \mathbb{X}$, and $1 \leq i \neq j \leq n$:*

$$p_{ij} \in [c_1, 1 - c_1], \quad p_{ij}^{(1,0,0)},\ p_{ij}^{(0,1,0)} \in [c_2, 1 - c_2], \text{ and}$$
$$\left| p_{ij}^{(r_1,r_2,r_3)} \right| \leq c_3 \text{ for any } 1 \leq r_1 + r_2 + r_3 \leq 3.$$

Assumption 4 lower-bounds the link probability and the first derivatives with respect to the individual effects, while uniformly upper-bounding all derivatives up to total order three. In conjunction with Assumption 2, it is satisfied by common TU/NTU specifications with smooth link functions (e.g., the cumulative distribution function of the logistic or the normal). This assumption is similar to Fernández-Val and Weidner (2016, Assumption 4.3(v)), which regulates the smoothness of the likelihood functions.[3]

Remark 3 (Role of symmetry in the proofs)**.** Our theoretical results are derived under the general symmetric link function $p(\alpha_i, \alpha_j, x_{ij}^\top \beta)$ in Assumption 4, without distinguishing between TU and NTU. The key structural property exploited throughout the proofs is symmetry in the first two arguments (i.e., $p(\alpha_i, \alpha_j, t) = p(\alpha_j, \alpha_i, t)$), which holds under both TU (where p depends on $\alpha_i + \alpha_j$) and NTU (where bilateral consent ensures the linking probability is symmetric). Thus, we do not rely on TU nor specific link function to prove the theory.

[3]The $p_{ij} \in [c_1, 1-c_1]$ bound rules out, in population, nodes with zero or unit linking probabilities. In finite samples, a node i may have degree $d_i = 0$ (isolated) or $d_i = n - 1$ (fully connected), causing the fixed-point iteration (2) to push $\widehat{\alpha}_i$ to $\pm\infty$. We recommend removing such nodes before estimation. Under the dense-network assumption, the probability of any node being isolated or fully connected vanishes exponentially in n, so this trimming does not alter the asymptotic theory.

3.1 JMM

Recall that the JMM module gives $(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})$ to start with. The next lemma concerns the existence and uniqueness of $\widehat{\boldsymbol{\alpha}}(\beta)$, as well as the convergence of $\boldsymbol{\alpha}^k(\beta)$ to $\widehat{\boldsymbol{\alpha}}(\beta)$ via (2).

Lemma 1. *If Assumptions 1, 2, and 4 hold, then there exists a unique $\widehat{\boldsymbol{\alpha}}(\beta)$ w.p.a.1 for each $\beta \in \{\beta \in \mathbb{B} | \, \|\beta - \beta_0\|_2 < c\}$ in a neighborhood around β_0, where $c > 0$ is small but fixed. Moreover, uniformly across all k, we have*

$$\|\boldsymbol{\alpha}^{k+2}(\beta) - \widehat{\boldsymbol{\alpha}}(\beta)\|_1 \leq \delta \|\boldsymbol{\alpha}^k(\beta) - \widehat{\boldsymbol{\alpha}}(\beta)\|_1,$$

for some fixed constant $\delta \in (0, 1)$.

Lemma 1 guarantees that $\widehat{\boldsymbol{\alpha}}(\beta) = \lim_{k\to\infty} \boldsymbol{\alpha}^k(\beta)$ and that the ℓ_1-distance between $\widehat{\boldsymbol{\alpha}}(\beta)$ and $\boldsymbol{\alpha}^k(\beta)$ decreases geometrically after every two iterations. Computing $\widehat{\boldsymbol{\alpha}}(\beta)$ is fast in the simulations, which is another advantage of our iterative algorithm. It is worth mentioning that we deviate from the existing methods (e.g., Theorem 1.5 of Chatterjee, Diaconis, and Sly, 2011 or the fixed point equation (17) of Graham, 2017) in this step by not requiring p_{ij} to be logistic or the link formation process to be TU. Instead, we use a gradient-descent-type iterative algorithm (2) to compute $\widehat{\boldsymbol{\alpha}}(\beta)$ as a function of β and show that it is a contraction mapping. As a result, it can accommodate general non-logistic link functions and NTU.

Although $\widehat{\boldsymbol{\alpha}}(\beta)$ is unique by Lemma 1 for any β that lies within a distance of c of β_0, in principle there could be multiple solutions to $m_2(\widehat{\boldsymbol{\alpha}}(\beta), \beta) = 0$. The next identification condition guarantees that any such $\widehat{\beta}$ is consistent for β_0. To state the assumption, we define the concentrated moment equation for β as

$$\bar{S}_n(\beta) := N^{-1}\mathbb{E}[m_2(\boldsymbol{\alpha}(\beta), \beta) | \mathbf{x}, \boldsymbol{\alpha}_0], \tag{6}$$

where $\boldsymbol{\alpha}(\beta)$ is the unique solution to $\mathbb{E}[\mathbf{m}_1(\boldsymbol{\alpha}, \beta) | \mathbf{x}, \boldsymbol{\alpha}_0] = \mathbf{0}_n$, a result from the proof of Lemma 1.

Assumption 5 (Identification of β_0)**.** *Suppose for all $\delta > 0$ and for n large enough,*

$$\inf_{\beta \in \mathbb{B}: \|\beta - \beta_0\|_2 \geq \delta} \left\|\bar{S}_n(\beta)\right\|_2 > 0 \text{ and } (\partial \bar{S}_n(\beta) / \partial \beta^\top)|_{\beta = \beta_0} \text{ has full rank.}$$

Assumption 5 identifies the low-dimensional parameter β_0, as discussed in Chen, Chernozhukov, Lee, and Newey (2014) for nonlinear models with high-dimensional nuisance parameters. The first part is the standard "unique minimizer" condition that makes β_0 the unique zero of $\bar{S}_n$ (van der Vaart, 2000, Page 45); the second part is a local rank condition—equivalently, the concentrated Jacobian $\mathrm{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12}$ is nonsingular at the truth. Substantively, Assumption 5 rules out cancellations in which the effect of β on the x-weighted link residuals can be exactly offset by adjusting the node fixed effects while preserving all degree moments.

Assumption 5 is the dyadic-network analogue of the within-variation condition in panel models, but residualization is against *additive node effects* rather than within-individual means. Let $e_i \in \mathbb{R}^n$ denote the ith standard basis vector, D the $N \times n$ dyad-incidence matrix whose row corresponding to dyad (i, j) equals $e_i + e_j$, X the $N \times K$ matrix stacking $x_{ij}^\top$ as rows, and $W_0 = \mathrm{diag}\{p_{ij,0}(1-p_{ij,0})\}$. Under the additive TU-logit specification $p_{ij} = \Lambda(\alpha_i + \alpha_j + x_{ij}^\top\beta)$, differentiating the profiled moment at β_0 gives

$$\left.\frac{\partial \bar{S}_n(\beta)}{\partial \beta^\top}\right|_{\beta_0} = -N^{-1}(M_D X)^\top W_0\,(M_D X), \qquad \text{where} \;\; M_D = I - D(D^\top W_0 D)^{-1} D^\top W_0,$$

so the rank condition holds iff X has nontrivial component orthogonal to the additive-node subspace $\mathrm{col}(D)$. Non-additive covariates such as $|x_i - x_j|$ or $x_i \cdot x_j$ pass this test trivially; $x_{ij} = x_i + x_j$ is the canonical violation, since β then absorbs into relabeled fixed effects $\widetilde{\alpha}_i \coloneqq \alpha_i + x_i\beta$. The global single-zero requirement is automatically satisfied under TU-logit from strict concavity of the population log-likelihood. Under NTU, where this concavity is unavailable, it reduces to parametric identifiability of the map $(\boldsymbol{\alpha}, \beta) \mapsto \{F(\alpha_i + x_{ij}\beta)F(\alpha_j + x_{ij}\beta)\}_{i<j}$, which holds under the same non-additivity. As a concrete check, in a two-group NTU-logit model with $x_{ij} = 1$ for cross-group dyads and $x_{ij} = 0$ within-group, the within-group subgraph pins down $\boldsymbol{\alpha}_0$ via $\Lambda(\alpha_i)\Lambda(\alpha_j)$ and any cross-group dyad then identifies β_0 by strict monotonicity of Λ.

In the next theorem, we prove that $\widehat{\beta}$ is consistent for β_0 and that $\widehat{\boldsymbol{\alpha}}$ is uniformly consistent for $\boldsymbol{\alpha}_0$ in the sup norm. Furthermore, we establish asymptotic normality for $\widehat{\beta}$. To state the result, define the concentrated Jacobian $\mathrm{J}_n(\boldsymbol{\alpha}, \beta)$ in (12) and its probability limit $\mathrm{J}_0 \coloneqq \mathrm{plim}_{n\to\infty} N^{-1}\mathrm{J}_{n,0}$. The bias vector $B_0 = (B_{10}, \ldots, B_{K0})^\top$ is given in (15), and the asymptotic sandwich variance Ω_0 is given in (16); both are defined in Appendix A. Lemma 5 in Appendix B establishes the existence and positive-definiteness of these limits.

Theorem 1. *If Assumptions 1–5 hold, then*

$$\widehat{\beta} \xrightarrow{p} \beta_0 \quad \text{and} \quad \|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|_\infty \xrightarrow{p} 0.$$

Furthermore, we have

$$\sqrt{N}(\widehat{\beta} - \beta_0) - \mathrm{J}_0^{-1} B_0 \xrightarrow{d} \mathcal{N}(0, \Omega_0).$$

Theorem 1 shows that $\widehat{\boldsymbol{\alpha}}$ is uniformly consistent for $\boldsymbol{\alpha}_{\mathbf{0}}$ and the JMM estimator $\widehat{\beta}$ is asymptotically normal. However, the limiting distribution of $\widehat{\beta}$ does not center around β_0. The bias term $\mathrm{J}_0^{-1} B_0$ arises from estimating $\boldsymbol{\alpha}_0$. The incidental parameter problem is common in the literature on nonlinear panel fixed effects regression with large N and T. Moreover, as Ω_0 is generally greater than ${\mathrm{I}_0}^{-1}$ in the semi-definiteness sense, $\widehat{\beta}$ does not achieve the CRLB. We refine $\widehat{\beta}$ by the following modules.

3.2 One-Step Estimator

The first refinement on $\widehat{\beta}$ concerns achieving the CRLB. We follow Le Cam (1969)'s one-step adjustment as specified in Module 2. Algebra shows

$$\mathbb{E}\left[\frac{\partial s_n(\boldsymbol{\alpha}, \beta_0)}{\partial \boldsymbol{\alpha}} \middle| \mathbf{x}, \boldsymbol{\alpha}_0\right] = \mathbf{0}_{K\times n} \text{ and } \mathbb{E}\left[\frac{\partial s_n(\boldsymbol{\alpha}, \beta_0)}{\partial \beta} \middle| \mathbf{x}, \boldsymbol{\alpha}_0\right] = -\mathrm{I}_n(\boldsymbol{\alpha}_{\mathbf{0}}, \beta_0).$$

Therefore, a Taylor expansion on the right-hand side of (5) yields

$$\widehat{\beta}_{\mathrm{OS}} - \beta_0 \approx \mathrm{I}_{n,0}^{-1} s_{n,0} \tag{7}$$

in large samples, where s_n and I_n are the concentrated score and information matrix defined in (4).

Remark 4 (OS and non-concavity in the fixed effects)**.** The normalizer in (5) is the outer-product Fisher information $\mathbf{I}_{11} = \mathbb{E}[\mathbf{s}_1 \mathbf{s}_1^\top \mid \mathbf{x}, \boldsymbol{\alpha}]$, not the negative concentrated Hessian $-\partial^2 \ell_n / \partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top$. The two agree in expectation at the truth by the information identity, but only $\mathbf{I}_{11}$ is positive semidefinite for every $(\boldsymbol{\alpha}, \beta)$ by construction; its invertibility and diagonal approximation (Lemma 3) rest on the positive per-dyad Fisher information guaranteed by Assumption 4, not on concavity of ℓ_n in $\boldsymbol{\alpha}$. Module OS therefore remains well-defined and first-order efficient under global non-concavity in the fixed effects, as arises structurally under NTU and under TU whenever the shock density is not log-concave (e.g., Student's t). A Newton update using $-\partial^2 \ell_n / \partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top$ would be ill-conditioned or undefined here.

To establish (7) rigorously and hence the asymptotic normality of $\widehat{\beta}_{\text{OS}}$, we impose an additional assumption on the conditioning of the information matrix (3). Define $\mathbf{D}(\boldsymbol{\alpha}, \beta) \coloneqq \text{diag}(\mathbf{I}_{11}(\boldsymbol{\alpha}, \beta))$ and $\mathbf{Q}(\boldsymbol{\alpha}, \beta) \coloneqq \mathbf{D}(\boldsymbol{\alpha}, \beta)^{-1/2}\mathbf{I}_{11}(\boldsymbol{\alpha}, \beta)\mathbf{D}(\boldsymbol{\alpha}, \beta)^{-1/2}$.

Assumption 6 (Information Matrix Conditioning)**.**

$$\sup_{(\boldsymbol{\alpha},\beta)\in\mathbb{A}\times\mathbb{B}} \left\|\mathbf{Q}(\boldsymbol{\alpha}, \beta)^{-1}\right\|_1 < c \textit{ for some positive constant } c.$$

The concentrated score s_n is constructed by projecting the fixed-effect scores out of the β score. Define the profiling weights

$$w_{ki}(\boldsymbol{\alpha}, \beta) \coloneqq \left[\mathbf{I}_{12}(\boldsymbol{\alpha}, \beta)^\top \mathbf{I}_{11}(\boldsymbol{\alpha}, \beta)^{-1}\right]_{ki}, \tag{8}$$

which determine how much of node i's fixed-effect score is subtracted from the raw score for β_k. In the additive TU-logit benchmark, the concentrated score takes the form $s_{n,k}(\boldsymbol{\alpha}, \beta) = \sum_{i<j}[y_{ij} - p_{ij}(\boldsymbol{\alpha}, \beta)](x_{ij,k} - w_{ki} - w_{kj})$, where $w_{ki} + w_{kj}$ is the best weighted least-squares approximation to $x_{ij,k}$ with additive node terms, a Frisch–Waugh–Lovell residualization. Assumption 6 is a stability condition on this projection: the standardized matrix $\mathbf{Q}$ normalizes $\mathbf{I}_{11}$ to unit diagonal, and bounding $\|\mathbf{Q}^{-1}\|_1$ prevents the node-level information contributions from becoming near-collinear as n grows. For additive models $p(\alpha_i, \alpha_j, x_{ij}^\top\beta) = F(\alpha_i + \alpha_j + x_{ij}^\top\beta)$ with smooth F, $p^{(1,0,0)} = p^{(0,1,0)} = F'$, which makes $\mathbf{I}_{11}$ diagonally dominant with off-diagonal entries of order $O(1)$. In this case, $\mathbf{I}_{11}(\boldsymbol{\alpha}, \beta)^{-1}$ can be approximated by $\mathbf{D}(\boldsymbol{\alpha}, \beta)^{-1}$ with maximum entrywise error $O(n^{-2})$ by Lemma 2, then Assumption 6 is not needed (see Remark A.1 in the Supplemental Material).

Remark 5. Lemma 4 in Appendix B shows that Assumption 6, together with Assumptions 1–4, implies the order bounds on w_{ki} and its derivatives that are used in the proof of Theorem 2: $\sup_{k,i}|w_{ki}| = O(1)$, $\sup_{k,i}\|\partial w_{ki}/\partial\beta\| = O(1)$, $\sup_{k,i}|\partial w_{ki}/\partial\alpha_i| = O(1)$, and $\sup_{k,i\neq j}|\partial w_{ki}/\partial\alpha_j| = O(n^{-1})$. Intuitively, the last bound says that no single foreign node j has order-1 leverage on node i's profiling weight, which is natural in dense networks where each node averages information over $n-1$ dyads.

The next theorem establishes the limit distribution of $\widehat{\beta}_{\text{OS}}$. Define the concentrated information limit

$$\mathrm{I}_0 \coloneqq \operatorname{plim}_{n\to\infty} N^{-1}\mathrm{I}_{n,0} \tag{9}$$

and the asymptotic bias $b_0 = (b_{10}, \ldots, b_{K0})^\top$ with

$$b_{k0} = \lim_{n\to\infty} \frac{1}{\sqrt{N}} \mathrm{Tr}[\mathbf{J}_{11,0}^{-1}\mathrm{Cov}(\mathbf{m}_{1,0}, \mathbf{s}_{1,0})\mathbf{W}_{k,0}], \quad k = 1, \ldots, K,$$

where $[\mathbf{W}_k(\boldsymbol{\alpha}, \beta)]_{ij} = \frac{\partial w_{ki}(\boldsymbol{\alpha},\beta)}{\partial \alpha_j}$ and the comma-zero subscript denotes evaluation at $(\boldsymbol{\alpha}_0, \beta_0)$.

Theorem 2. *If Assumptions 1–6 hold, then*

$$\sqrt{N}(\widehat{\beta}_{\mathrm{OS}} - \beta_0) - \mathrm{I}_0^{-1} b_0 \xrightarrow{d} \mathcal{N}(0, \mathrm{I}_0^{-1}).$$

Theorem 2 shows $\widehat{\beta}_{\mathrm{OS}}$ achieves the CRLB asymptotically. In the proof of Theorem 2, we show that b_0 is $O(1)$ and depends on the covariance matrix between $\mathbf{m}_1$ and $\mathbf{s}_1$. This is because our plug-in estimator for $\boldsymbol{\alpha}$ is obtained from the moment estimating equation $\mathbf{m}_1$, and the one-step estimator (5) uses information from $\mathbf{s}_1$ to concentrate out $\boldsymbol{\alpha}$. As a result, the covariance between $\mathbf{m}_1$ and $\mathbf{s}_1$ determines the magnitude of the term b_0 in the asymptotic bias of Theorem 2.

3.3 Bagging

While reaching the CRLB, Theorem 2 reveals that $\widehat{\beta}_{\mathrm{OS}}$ incurs an asymptotic bias. As discussed in Module 3, one way to debias $\widehat{\beta}_{\mathrm{OS}}$ is to use split-network jackknife to self-cancel the leading bias. However, it doubles the asymptotic variance, making the confidence interval wider and less informative. Additionally, it suffers from computational instability because the network is split only once by random. The solution we propose is to split the network randomly many times, compute the split-network jackknife estimator from each split, then average them up. This is equivalent to bagging (Breiman, 1996) on a split-network jackknife estimator.

To motivate the bagging method, in theory there are a total of $T_n := \binom{n}{n/2}$ possible ways to divide the network. However, T_n can be very large for a moderate sample size n. For example, $n = 100$ produces $T_n = \binom{100}{50} \simeq 1.009 \times 10^{29}$, which is an astronomical number. We solve this problem by choosing $\widetilde{T}_n \ll T_n$ in the BG module. In the simulations, we set $\widetilde{T}_n = 2n$ and find that the results are robust to this choice.

The next theorem shows that when n and $\widetilde{T}_n$ go to infinity, $\widehat{\beta}_{\mathrm{BG}}$ is asymptotically normal, unbiased, and efficient.

Theorem 3. *If Assumptions 1–6 hold, then*

$$\sqrt{N}(\widehat{\beta}_{\mathrm{BG}} - \beta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{I}_0^{-1})$$

as $n \to \infty$ and $\widetilde{T}_n \to \infty$.

Theorem 3 is the main theoretical result of this paper. A few remarks are in order to discuss its implications and connections with the literature. First, $\widehat{\beta}_{\mathrm{BG}}$ involves three modules—JMM, OS, and BG—which play different roles. Module JMM provides an initial consistent yet biased estimator, which is fed into Module OS to achieve the CRLB. Module BG corrects for the bias in the OS estimator via split-network jackknife while maintaining the efficiency through bagging.

Second, a similar idea to $\widehat{\beta}_{\mathrm{OS-SJ}}$ in a panel setting with fixed effects is presented in Dhaene and Jochmans (2015). Although related, $\widehat{\beta}_{\mathrm{BG}}$ is preferred over $\widehat{\beta}_{\mathrm{OS-SJ}}$ because $\widehat{\beta}_{\mathrm{OS-SJ}}$ has an asymptotic variance of $2\mathrm{I}_0^{-1}$ while $\widehat{\beta}_{\mathrm{BG}}$'s is I_0^{-1}.

Third, one may be inclined to apply BG to the initial JMM estimator directly and bypass the one-step approximation. Indeed, BG can correct for the asymptotic bias of the JMM estimator. However, the JMM-BG estimator is not efficient because the asymptotic variance of the initial JMM estimator is preserved through bagging.

Finally, sample splitting across individuals introduces a degree of extra randomness, which motivates Fernández-Val and Weidner (2016, Footnote 8) to suggest averaging of all possible T_n partitions and point out that the average over $\widetilde{T}_n \ll T_n$ splits is sufficient. The BG estimator in our context not only eliminates randomness from sample splitting but also simultaneously achieves efficiency and bias correction. Furthermore, our Theorem 3 provides formal asymptotic results to justify the use of the BG estimator.

Remark 6 (Multiple networks)**.** The estimator extends to the multi-network setting. Consider V independent networks indexed by $v = 1, \dots, V$ (V fixed) sharing a common slope β_0 but with network-specific node fixed effects:

$$\Pr(Y_{ij,v} = 1 \mid X_{ij,v}, \alpha_{i0,v}, \alpha_{j0,v}) = p\big(\alpha_{i0,v}, \alpha_{j0,v}, X_{ij,v}^{\top}\beta_0\big), \quad 1 \le i \ne j \le n_v,\ 1 \le v \le V.$$

The estimator pools across networks at the β level while keeping $\boldsymbol{\alpha}_v$ network-specific: the JMM module solves $\widehat{\boldsymbol{\alpha}}_v(\beta)$ within each network from its own degree equations, and $\widehat{\beta}$ solves the pooled moment $\sum_v m_{2,v}(\widehat{\boldsymbol{\alpha}}_v(\beta), \beta) = 0$; the one-step refinement uses the pooled concentrated information $\sum_v \mathrm{I}_{n_v}(\widehat{\boldsymbol{\alpha}}_v, \widehat{\beta})$ and score $\sum_v s_{n_v}(\widehat{\boldsymbol{\alpha}}_v, \widehat{\beta})$; bagging is applied network-by-network (each $\mathcal{I}_{n_v}$ is split independently). Assumptions 1–4 and 6 hold per network, and Assumption 5 is imposed on the pooled $\bar{S}_n$. Theorems 1–3 extend with $N = \sum_v \binom{n_v}{2}$ as $\min_v n_v \to \infty$ (Proposition A.1 in the Supplement).

4 Extensions

This section extends our framework in two directions: average partial effects for the binary outcome, and the behavior of our estimators when the link function $p(\cdot)$ is misspecified.

Average partial effects. Beyond the homophily parameters themselves, researchers and policy makers may also be interested in population averages over the distribution of exogenous regressors and fixed effects. One leading example concerns the conditional mean of the outcome given covariates and fixed effects

$$\mathbb{E}\left[Y_{ij} \middle| X_{ij} = x_{ij}, \boldsymbol{\alpha}\right] = p_{ij}(\boldsymbol{\alpha}, \beta_0). \tag{10}$$

Here, the partial effects are defined as the differences or derivatives of (10) with respect to components of X_{ij}, say $X_{ij,k}$, the kth coordinate of X_{ij}. We suppress its dependence on Y and X and define the partial effect of $x_{ij,k}$ for the dyad (i,j) as

$$\Delta_{ij,k}(\alpha_i, \alpha_j, \beta) = \begin{cases} p(\alpha_i, \alpha_j, \beta_k + x_{ij,-k}^\top \beta_{-k}) - p(\alpha_i, \alpha_j, x_{ij,-k}^\top \beta_{-k}) & (b) \\ \beta_k p^{(0,0,1)}(\alpha_i, \alpha_j, x_{ij}^\top \beta) & (c) \end{cases}$$

where "(b)" corresponds to binary $x_{ij,k}$ while "(c)" refers to continuous $x_{ij,k}$. Define $\Delta_{ij} = (\Delta_{ij,1}, \ldots, \Delta_{ij,K})^\top$. Then, the unconditional APEs are

$$\delta_0 = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{n}\sum_{j>i}\Delta_{ij}(\alpha_i, \alpha_j, \beta_0)\right]. \tag{11}$$

Plugging the JMM estimator $(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})$ into (11) yields an estimator for the APEs

$$\widehat{\delta} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j>i}\Delta_{ij}(\widehat{\alpha}_i, \widehat{\alpha}_j, \widehat{\beta}).$$

Define an (infeasible) $\bar{\Delta}_n = \frac{1}{N}\sum_{i=1}^{n}\sum_{j>i}\Delta_{ij}(\alpha_{i0}, \alpha_{j0}, \beta_0)$. Let the split-jackknife estimator and the bagging estimator of the APE be

$$\widehat{\delta}_{\mathrm{SJ}} := 2\widehat{\delta} - \frac{1}{2}(\widehat{\delta}_1 + \widehat{\delta}_2) \quad \text{and} \quad \widehat{\delta}_{\mathrm{BG}} := \frac{1}{\widetilde{T}_n}\sum_{t=1}^{\widetilde{T}_n}\widehat{\delta}_{\mathrm{SJ}}^{(t)},$$

respectively. Here, $(\widehat{\delta}_1, \widehat{\delta}_2)$ are the plug-in estimators based on two sub-networks after a random split of the nodes and $\{\widehat{\delta}_{\mathrm{SJ}}^{(t)}\}_{t=1}^{\widetilde{T}_n}$ are split-network jackknife estimators based on $\widetilde{T}_n$ random splits. The next theorem shows that the bias of $\widehat{\delta}$ is asymptotically negligible. We use a central limit theorem for U-statistics (van der Vaart, 2000, Theorem 12.3) to prove it. To state the result precisely, we incorporate the asymptotically vanishing bias terms, as in Fernández-Val and Weidner (2016, Theorem 4.2), and establish that $\widehat{\delta}_{\mathrm{SJ}}$ and $\widehat{\delta}_{\mathrm{BG}}$ are equivalent to $\widehat{\delta}$ asymptotically. Section 5 presents numerical evidence that supports this claim.

To state the next theorem, we define $\Sigma_{\delta,n} := \frac{\Sigma_\Delta}{N} + \frac{4\Sigma_\delta}{n}$, where Σ_Δ is defined in (18) and $\Sigma_\delta = \mathrm{Cov}\left(\Delta_{12}(\alpha_1, \alpha_2, \beta_0), \Delta_{13}(\alpha_1, \alpha_3, \beta_0)\right)$.[4] The two bias terms B_β and B_α in the next theorem are defined in (17).

Theorem 4. *If Assumptions 1–5 hold and $\bar{\Delta}_n$ is a non-degenerate U-statistic, then*

$$\Sigma_{\delta,n}^{-1/2}\left(\widehat{\delta} - \delta_0 - \frac{1}{\sqrt{N}}B_\beta - \frac{1}{\sqrt{N}}B_\alpha\right) \xrightarrow{d} \mathcal{N}(0, I_K) \text{ and}$$
$$\Sigma_{\delta,n}^{-1/2}\left(\widehat{\delta}_{\mathrm{BG}} - \delta_0\right) \xrightarrow{d} \mathcal{N}(0, I_K).$$

In Theorem 4, the rate of convergence of $\widehat{\delta}$ (and $\widehat{\delta}_{\mathrm{BG}}$) is $\sqrt{n}$ instead of $\sqrt{N}$. The slower convergence rate in Theorem 4 makes the bias terms introduced by estimating $\boldsymbol{\alpha_0}$ asymptotically negligible. Note that B_β reflects the bias of the plug-in estimator $\widehat{\beta}$ whereas B_α arises from the incidental parameter bias of the plug-in estimator $\widehat{\boldsymbol{\alpha}}$.

For the components of $\Sigma_{\delta,n}$, Σ_Δ is the asymptotic variance of $\sqrt{N}(\widehat{\delta} - \bar{\Delta}_n)$ and $4\Sigma_\delta$ is the asymptotic variance of $\sqrt{n}(\bar{\Delta}_n - \delta_0)$. Σ_δ can be estimated by

$$\widehat{\Sigma}_\delta = \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} u_{ij} u_{ik}^\top,$$

where $u_{ij} := \Delta_{ij}(\widehat{\alpha}_i, \widehat{\alpha}_j, \widehat{\beta}) - \widehat{\delta}$ and $\sum_{i\neq j\neq k}$ denotes the sum over all triads (i, j, k) with distinct indices. This symmetrized estimator is permutation-invariant in finite samples and is consistent by the law of large numbers for U-statistics. Although the variance term Σ_Δ/N is dominated asymptotically by $4\Sigma_\delta/n$ in Theorem 4, we find in simulations that including it improves the coverage probabilities. This is particularly relevant when unobserved heterogeneity is limited (i.e., $\alpha_i \approx \alpha_j$ for most nodes), in which case the U-statistic leading term Σ_δ may be small and the higher-order term

[4] We acknowledge the presence of high-order variance terms arising from U-statistics, but omit these terms for simplicity. Our simulation results confirm that this omission does not compromise robustness.

Σ_Δ/N contributes non-negligibly to the finite-sample variance. We therefore recommend including both terms in practice. If one instead targets $\bar{\Delta}_n$, the asymptotic result becomes $\sqrt{N}(\widehat{\delta}-\bar{\Delta}_n)-B_\beta-B_\alpha \xrightarrow{d} \mathcal{N}(0,\Sigma_\Delta)$, generalizing Theorem 2 of Chen, Fernández-Val, and Weidner (2021) to our setting.

Link function misspecification. The preceding results assume that the link function $p(\cdot)$ is correctly specified. In Section B of the Supplemental Material, we show that our estimators remain well-behaved when $p(\cdot)$ is replaced by a misspecified link function $q(\cdot)$. Specifically, the JMM estimator $\widehat{\beta}$ converges to a well-defined pseudo-true value β_{n*} that solves a pseudo-population moment equation, with a sandwich-form asymptotic variance that can be consistently estimated. The one-step and BG estimators center on a projected pseudo-true value $\beta_{n\sharp}$ that accounts for the nonzero population concentrated score under misspecification. Crucially, the BG estimator retains its bias-correction property: $\sqrt{N}(\widehat{\beta}_{\mathrm{BG}}-\beta_{n\sharp}) \xrightarrow{d} \mathcal{N}(0,\Gamma_*)$, where the sandwich covariance Γ_* reflects the discrepancy between the true and misspecified likelihoods. Section C of the Supplemental Material provides supporting numerical evidence.

5 Monte Carlo Simulations

We conduct Monte Carlo simulations to evaluate the finite-sample performance of our estimators under both the TU and NTU specifications. Additional exercises—the non-concavity challenge of direct MLE, fixed-effect recovery, APE estimation, link-function misspecification, sparser networks, and extended TU results—are reported in Section C of the Supplemental Material.

The data generating process (DGP) is as follows. We set $\beta_0 = (1,-1)^\top$, and draw the first covariate of X_{ij} as $X_{1,ij} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(0.3),\ X_{1,ij} = X_{1,ji}$. In this way, we allow for a discrete variable in X_{ij}. For the second covariate of X_{ij}, we draw $X_i \overset{\text{i.i.d.}}{\sim} U(-0.5,0.5)$ and let $X_{2,ij} = |X_i - X_j|$. Next, we generate the individual fixed effects by setting $\alpha_i = 0.75 \times \xi_i + 0.25 \times X_i$, where $\xi_i \overset{\text{i.i.d.}}{\sim} U(-0.5,0.5)$ and is independent of all other variables so that α_i and X_{ij} are correlated via X_i.

We consider two specifications for the link function. Under *TU*, a dyadic surplus determines the link:

$$Y_{ij}^{\mathrm{TU}} = \mathbb{1}\{\alpha_i + \alpha_j + X_{ij}^\top \beta_0 - \epsilon_{ij} > 0\},$$

with ϵ_{ij} drawn i.i.d. from either the standard logistic or standard normal distribution, yielding the logit and probit links respectively. Under *NTU*, link formation requires bilateral consent:

$$Y_{ij}^{\text{NTU}} = \mathbb{1}\{\alpha_i + X_{ij}^{\top}\beta_0 - \epsilon_{ij} > 0\} \cdot \mathbb{1}\{\alpha_j + X_{ij}^{\top}\beta_0 - \epsilon_{ji} > 0\},$$

with ϵ_{ij} drawn i.i.d. from the standard logistic distribution.

For all simulations in this paper, we run 1,000 replications. For the baseline results, we set $n = 100$ and 200, which is comparable to the size of the data used in our empirical applications.[5] We report the mean and median bias, standard deviation, mean and median absolute bias, and the root mean squared error (RMSE) across replications.

Table 1 reports the TU baseline results at $n = 100$ under both link functions. The BG estimator substantially reduces the bias of JMM and OS with essentially the same dispersion, achieving the lowest RMSE, and its coverage probabilities lie close to their nominal levels. The patterns are robust across logit and probit links, consistent with Theorem 3.

Table 2 reports the NTU results for $n = 100$ and 200 when the network has a density of 25%. Here are the main observations when $n = 100$. First, in terms of the bias, the BG estimator performs significantly better than JMM and OS, which is consistent with Theorem 3. Second, BG works very well in simultaneously achieving bias-correction and low standard deviation, leading to the lowest RMSE.[6] Third, the coverage probabilities of the BG confidence intervals are close to their nominal levels, while JMM and OS exhibit systematic undercoverage due to incidental parameter bias, precisely the problem that BG is designed to correct. Finally, the mean standard errors implied by the asymptotic theory are close to the standard deviations computed from the Monte Carlo simulations across all estimators. We also find that the quantiles of the empirical distributions for all estimators are well approximated by the same quantiles of the corresponding asymptotic normal distributions. These results further support our theoretical findings. When $n = 200$, the performance of all the estimators improves. The RMSEs, for example, are about half the size of those when $n = 100$, which is expected given the $\sqrt{N}$-convergence rate and $\sqrt{N} = O(n)$. The coverage

[5] We also run the same exercise for $n = 50$, and the conclusions remain largely the same.

[6] Though not reported in the tables, we find that SJ (without BG) doubles the variance of JMM and OS estimators in the simulations, which is in line with the theory.

Table 1: TU baseline estimation results for β_0 ($n = 100$)

Logit	JMM		OS		BG	
	β_1	β_2	β_1	β_2	β_1	β_2
Mean Bias	2.32	-1.97	2.44	-1.99	0.15	0.21
Median Bias	2.33	-2.25	2.43	-2.35	0.14	-0.20
Standard Deviation	6.96	14.67	6.97	14.74	6.79	14.46
Mean Standard Error	6.74	14.56	6.75	14.56	6.73	14.54
Mean Absolute Bias	5.82	11.87	5.87	11.93	5.39	11.54
Median Absolute Bias	4.87	9.98	4.90	10.09	4.52	9.90
RMSE	7.33	14.80	7.38	14.87	6.79	14.46
90% Coverage Rate	86.6	90.1	86.0	89.9	89.1	89.5
95% Coverage Rate	92.1	95.1	92.1	95.1	94.5	95.3
Probit	JMM		OS		BG	
	β_1	β_2	β_1	β_2	β_1	β_2
Mean Bias	1.95	-1.67	2.06	-1.68	-0.05	0.43
Median Bias	1.81	-1.88	1.93	-1.94	-0.21	0.32
Standard Deviation	4.44	9.46	4.44	9.50	4.32	9.35
Mean Standard Error	4.37	9.35	4.36	9.34	4.35	9.32
Mean Absolute Bias	3.88	7.77	3.92	7.80	3.49	7.53
Median Absolute Bias	3.41	6.75	3.40	6.83	2.90	6.51
RMSE	4.84	9.61	4.89	9.65	4.32	9.36
90% Coverage Rate	85.8	89.4	85.5	89.9	90.2	89.7
95% Coverage Rate	91.6	94.6	91.3	94.2	95.5	95.4

Note: All values have been multiplied by 100.

probabilities also improve.[7]

Finite-sample evidence for the APE estimator of Theorem 4 is reported in Section C of the Supplemental Material. The plug-in estimator exhibits near-zero bias and coverage close to the nominal level under both TU and NTU, consistent with the $O(n^{-1})$ bias bound. Extended results on fixed-effect recovery, link-function misspecification, and sparser networks are also provided there.

[7] We have checked the stability of $\widehat{\beta}_{\mathrm{BG}}$ for $\widetilde{T}_n$ chosen from a wide range, $\{50, 100, 200, 400, 600, 800\}$. The bias, RMSE and coverage are almost identical across different choices of $\widetilde{T}_n$.

Table 2: NTU baseline estimation results for β_0

$n = 100$	JMM		OS		BG	
	β_1	β_2	β_1	β_2	β_1	β_2
Mean Bias	2.95	-2.91	2.77	-2.71	-0.37	0.33
Median Bias	2.90	-3.01	2.77	-2.79	-0.40	0.24
Standard Deviation	5.71	13.05	5.67	13.02	5.51	12.69
Mean Standard Error	5.67	12.98	5.66	12.92	5.66	12.92
Mean Absolute Bias	5.17	10.68	5.07	10.61	4.45	10.12
Median Absolute Bias	4.35	8.89	4.21	8.92	3.92	8.53
RMSE	6.42	13.37	6.31	13.30	5.52	12.70
90% Coverage Rate	83.7	89.7	84.0	89.5	91.2	90.9
95% Coverage Rate	91.8	94.1	92.5	94.4	95.6	95.5
$n = 200$	JMM		OS		BG	
	β_1	β_2	β_1	β_2	β_1	β_2
Mean Bias	1.54	-1.71	1.47	-1.61	-0.06	-0.12
Median Bias	1.43	-1.73	1.33	-1.71	-0.20	-0.28
Standard Deviation	2.85	6.41	2.85	6.41	2.80	6.33
Mean Standard Error	2.78	6.35	2.78	6.32	2.78	6.32
Mean Absolute Bias	2.57	5.36	2.54	5.34	2.28	5.12
Median Absolute Bias	2.29	4.66	2.23	4.72	1.96	4.47
RMSE	3.24	6.63	3.20	6.61	2.80	6.33
90% Coverage Rate	84.8	88.5	85.3	88.0	90.2	90.0
95% Coverage Rate	90.8	93.9	90.9	94.3	95.3	95.0

Note: All values have been multiplied by 100.

6 Empirical Applications

This section presents two empirical applications that together exhibit the TU and NTU regimes, respectively. First, under TU, we apply our method to the Townsend Thai village networks (Kinnan, Samphantharak, Townsend, and Vera-Cossio, 2024), pooling across 16 villages to estimate link formation for financial, operations, and labor transactions. We find that kinship has a strong and broadly positive effect across all three networks, and that greater absolute net-worth differences between households are also associated with more frequent linking, a pattern consistent with gains-from-trade in transactional relationships under TU. Second, under NTU, we revisit the Nyakatoke risk-sharing network (De Weerdt, 2004), where our method indicates that wealth differences have no statistically significant effect on link formation, and we provide economic intuition for this result.

The two applications illustrate the framework's ability to accommodate both modeling regimes. For the Thai village networks, households transact in goods, labor, and financial instruments that carry transferable value, so we adopt the TU specification (de Paula, 2020b). For the Nyakatoke risk-sharing network, links record mutual acknowledgements without explicit transfers or side payments, so we adopt the NTU specification. In both applications, we take ϵ_{ij} to be standard logistic and report the JMM, OS, and BG estimates together with the plug-in and bagging APEs.

The BG estimators and standard errors in the two empirical applications are robust to different choices of $\widetilde{T}_n$, consistent with our simulation results. Though the empirical network densities are relatively low, our simulation results at comparable density (Section C of the Supplemental Material, 8.6%) suggest that the estimators remain well-behaved. Formal sparse-network guarantees are left for future work.

6.1 Townsend Thai Village Networks

We apply our method to the Townsend Thai monthly panel (Kinnan, Samphantharak, Townsend, and Vera-Cossio, 2024), covering 16 rural and peri-urban Thai villages with an average of 44 households per village (pooled sample of $N = 15{,}641$ dyads). We construct three binary network indicators from the monthly transaction records, aggregated to the annual level: a *financial* link (exchange of gifts or informal loans/repayments), an *operations* link (transactions in production inputs, intermediate goods, or output sales), and a *labor* link (hiring or labor exchange). Our dyadic covariates are (i) the (ln) demographic difference between two households, based on household composition and head characteristics; (ii) the (ln) net-worth difference; and (iii) an indicator for kinship. A detailed data description and summary statistics are provided in Section D.1 of the Supplemental Material.

To handle the multi-village structure, we estimate a pooled TU model with common slope β_0 and village-specific node fixed effects following Remark 6:

$$\Pr(Y_{ij,v} = 1 \mid X_{ij}, \alpha_{i0,v}, \alpha_{j0,v}) = F\big(\alpha_{i0,v} + \alpha_{j0,v} + X_{ij}^\top \beta_0\big), \quad 1 \le v \le V = 16,$$

where F is the logistic or normal CDF.

6.1.1 Year Selection and Estimation Results

Because our asymptotic theory assumes a dense-network regime, we select for each network type the year with the highest pooled density: year 3 for the financial network (1.46%), year 2 for the operations network (5.36%), and year 3 for the labor network (7.79%). Table 3 reports the pooled BG estimates and APEs under both the logit and probit link functions.

Table 3: BG estimates and APEs for the Townsend Thai village networks

	Coefficients		APEs	
Variables	Logit	Probit	Logit	Probit
Financial network				
(ln) Demographic difference	0.2629	0.1249	0.0029	0.0024
	(0.1862)	(0.0967)	(0.0017)	(0.0016)
(ln) Net-worth difference	0.2567	0.1293	0.0028	0.0024
	(0.1059)	(0.0632)	(0.0011)	(0.0010)
Kinship	3.2259	1.9318	0.0773	0.0896
	(0.2676)	(0.1512)	(0.0097)	(0.0101)
Operations network				
(ln) Demographic difference	0.1553	0.1006	0.0050	0.0061
	(0.1021)	(0.0532)	(0.0029)	(0.0029)
(ln) Net-worth difference	0.1221	0.0658	0.0040	0.0040
	(0.0563)	(0.0294)	(0.0016)	(0.0016)
Kinship	1.9651	1.0820	0.0973	0.0998
	(0.1778)	(0.0964)	(0.0119)	(0.0121)
Labor network				
(ln) Demographic difference	0.0569	0.0316	0.0029	0.0029
	(0.0771)	(0.0418)	(0.0035)	(0.0035)
(ln) Net-worth difference	0.1691	0.1056	0.0085	0.0096
	(0.0456)	(0.0251)	(0.0022)	(0.0022)
Kinship	2.2562	1.3040	0.1711	0.1774
	(0.1532)	(0.0856)	(0.0143)	(0.0147)

Note: Pooled BG estimates of the common slope β_0 in the TU specification with village-specific node fixed effects, using the logit and probit link functions. For each network type, we select the year with the highest pooled density to stay close to the dense-network regime assumed by the theory. Standard errors in parentheses.

Three findings stand out. First, *kinship* is a strong and highly significant predictor of link formation across all three networks, with the largest APE in the labor network (an increase in linking probability of about 17 percentage points). This is consistent with Kinnan, Samphantharak, Townsend, and Vera-Cossio (2024), who document that kinship ties are persistent predictors of transactions in these villages over a long time horizon. Second, the coefficient on *(ln) net-worth difference* is positive and statistically significant in all three networks, indicating that households with larger wealth disparities transact more frequently under the TU regime. Economically, this pattern is consistent with gains-from-trade in dyadic transactions: wealthier households can supply credit, labor demand, or production inputs that poorer households consume or provide, yielding bilateral surplus that the TU framework predicts. The contrast with the Nyakatoke finding below, where absolute wealth differences have no significant effect under NTU, highlights the role of the modeling regime: when linking requires mutual consent without transfers, the gains-from-trade channel is muted. Third, the coefficient on *(ln) demographic difference* is positive (directionally consistent with heterophily: demographically distant households transacting more often) but statistically insignificant, suggesting that household composition is not a primary driver of transactional linking after controlling for kinship and wealth. The logit and probit specifications yield qualitatively identical patterns, and the implied APEs are nearly identical in sign and magnitude.

6.2 Nyakatoke Risk-Sharing Network

We apply our method to the Nyakatoke risk-sharing network in Tanzania (De Weerdt, 2004), which covers 114 households with $N = 6{,}441$ dyadic observations. We estimate the model using three covariates: absolute (ln) wealth difference, (ln) distance, and a categorical kinship/religion tie variable. A detailed data description and summary statistics are provided in Section D.2 of the Supplemental Material.

6.2.1 Results and Discussion

Table 4 presents the estimation results for the homophily coefficients and the APEs. The estimated coefficient on wealth difference is negative under all three methods, but the null of zero cannot be rejected under the BG asymptotic distribution. The NTU bilateral-consent logic provides a natural interpretation: similar-wealth pairs lack the capacity to insure each other against large shocks, while substantially unequal pairs face veto by the wealthier household whose expected surplus from the arrangement is typically negative. Both effects push the average wealth-difference effect toward zero. Gao, Li, and Xu (2023) also obtain a negative coefficient but without inference; our framework complements theirs by quantifying statistical uncertainty through the parametric link function.

In addition to the wealth difference, under BG the coefficient for *distance* is significantly negative at -0.8098, and that of *tie* is significantly positive at 0.5875. The results are intuitive. We further report the APEs in the last two columns of Table 4. We find that the APE of wealth difference is not significant based on either the plug-in or the bagging estimator. Distances between households and social ties, on the other hand, matter more significantly in terms of the APE. The distribution of the estimated individual fixed effects $\widehat{\boldsymbol{\alpha}}$ is reported in Figure C.3 of the Supplemental Material.

Table 4: Estimation results for the Nyakatoke network

Variables	Coefficients			APEs	
	JMM	OS	BG	Plug-in	Bagging
(ln) wealth difference	-0.0882	-0.0974	-0.0783	-0.0065	-0.0096
	(0.0680)	(0.0641)	(0.0641)	(0.0052)	(0.0052)
(ln) distance	-0.7824	-0.8636	-0.8098	-0.0576	-0.0638
	(0.0530)	(0.0536)	(0.0536)	(0.0066)	(0.0064)
Tie	0.6714	0.6287	0.5875	0.0514	0.0500
	(0.0548)	(0.0556)	(0.0556)	(0.0060)	(0.0059)

Note: Standard errors are reported in the parentheses.

7 Conclusion

This paper develops a unified estimation and inference framework for dyadic network formation models with individual fixed effects under both transferable and non-transferable utilities. The key methodological innovation is a bagging estimator that combines a one-step approximation to the MLE with a split-network jackknife, delivering asymptotically unbiased and efficient inference for the homophily parameters without requiring concavity of the log-likelihood in high-dimensional fixed effects or distributional assumptions on the link function. As by-products, we obtain ℓ_∞-consistent estimators of the individual fixed effects and asymptotically normal estimators of the average partial effects. Two empirical applications, the Townsend Thai village networks (TU) and the Nyakatoke risk-sharing network (NTU), illustrate that the method is straightforward to implement under both modeling regimes and yields interpretable results.

Several extensions merit future research. First, extending the framework to *directed networks* would capture asymmetric relationships such as trade flows or citations by introducing separate in- and out-degree fixed effects. Second, establishing *sparse-network* asymptotics, where the average degree grows slower than n, would relax the dense-network regime assumed by our theory and broaden the method's applicability. Third, incorporating *strategic interactions*, where link decisions depend on others' anticipated choices, would bring the framework closer to equilibrium models of network formation.

Appendix

A Notation and Definitions

We collect additional matrix definitions used in the main text and appendix.

Jacobian of the moment equations. The Jacobian matrix $\mathbf{J}(\boldsymbol{\alpha}, \beta) \coloneqq \nabla \mathbf{m}(\boldsymbol{\alpha}, \beta)$ is partitioned as

$$\mathbf{J}(\boldsymbol{\alpha}, \beta) = \begin{pmatrix} \mathbf{J}_{11}(\boldsymbol{\alpha}, \beta) & \mathbf{J}_{12}(\boldsymbol{\alpha}, \beta) \\ \mathbf{J}_{21}(\boldsymbol{\alpha}, \beta) & \mathrm{J}_{22}(\boldsymbol{\alpha}, \beta) \end{pmatrix},$$

where $\mathbf{J}_{11}$ is $n \times n$ with $[\mathbf{J}_{11}]_{ij} = -p_{ij}^{(0,1,0)}$ for $i \neq j$ and $[\mathbf{J}_{11}]_{ii} = -\sum_{j \neq i} p_{ij}^{(1,0,0)}$; the ith row of $\mathbf{J}_{12}$ $(n \times K)$ is $-\sum_{j \neq i} p_{ij}^{(0,0,1)} x_{ij}^\top$; the ith column of $\mathbf{J}_{21}$ $(K \times n)$ is $-\sum_{j \neq i} p_{ij}^{(1,0,0)} x_{ij}$; and $\mathrm{J}_{22} = -\sum_{i<j} p_{ij}^{(0,0,1)} x_{ij} x_{ij}^\top$. The concentrated Jacobian for β is

$$\mathrm{J}_n(\boldsymbol{\alpha}, \beta) \coloneqq \mathrm{J}_{22} - \mathbf{J}_{21} \mathbf{J}_{11}^{-1} \mathbf{J}_{12}. \tag{12}$$

Variance of the moment equations. Define the variance blocks $\mathbf{V}_{11}(\boldsymbol{\alpha}, \beta)$ $(n \times n)$, $\mathbf{V}_{12}(\boldsymbol{\alpha}, \beta)$ $(n \times K)$, and $\mathrm{V}_{22}(\boldsymbol{\alpha}, \beta)$ $(K \times K)$ by $[\mathbf{V}_{11}]_{ij} = p_{ij}(1 - p_{ij})$ for $i \neq j$ and $[\mathbf{V}_{11}]_{ii} = \sum_{j \neq i} p_{ij}(1 - p_{ij})$; the ith row of $\mathbf{V}_{12}$ is $\sum_{j \neq i} p_{ij}(1 - p_{ij}) x_{ij}^\top$; and $\mathrm{V}_{22} = \sum_{i<j} p_{ij}(1 - p_{ij}) x_{ij} x_{ij}^\top$.

Hessian of the log-likelihood. The Hessian blocks $\mathbf{H}_{11}(\boldsymbol{\alpha},\beta)$, $\mathbf{H}_{12}(\boldsymbol{\alpha},\beta)$, and $\mathrm{H}_{22}(\boldsymbol{\alpha},\beta)$ are the sub-blocks of $\nabla^2\ell_n(\boldsymbol{\alpha},\beta)$ partitioned conformably with $(\boldsymbol{\alpha},\beta)$.

Bias and variance of $\widehat{\beta}$. For each $i\in\mathcal{I}_n$, define the symmetric matrix $\mathbf{M}_{i,n}(\boldsymbol{\alpha},\beta)$ with entries

$$\begin{aligned}&(\mathbf{M}_{i,n})_{ij} = -(p_{ij}^{(1,1,0)} + p_{ji}^{(2,0,0)})\ (j\neq i),\ (\mathbf{M}_{i,n})_{ii} = -\sum_{j\neq i} p_{ij}^{(2,0,0)},\\ &(\mathbf{M}_{i,n})_{ll} = -p_{li}^{(2,0,0)}\ (l\neq i),\ \text{and}\ (\mathbf{M}_{i,n})_{ab}=0\ \text{otherwise.}\end{aligned} \tag{13}$$

Let $\varphi_k(\boldsymbol{\alpha},\beta)^\top := e_k^\top \mathbf{J}_{21}\mathbf{J}_{11}^{-1}$ and $\mathbf{G}_k(\boldsymbol{\alpha},\beta) := \sum_{i=1}^n \varphi_{k,i}\mathbf{M}_{i,n}$, and define $\mathbf{R}_k(\boldsymbol{\alpha},\beta)$ with entries

$$\begin{aligned}(\mathbf{R}_k)_{ij} &= p_{ij}^{(1,1,0)} x_{ij,k},\ 1\le i\neq j\le n,\\ (\mathbf{R}_k)_{ii} &= \sum_{j\neq i} p_{ij}^{(2,0,0)} x_{ij,k},\ i\in\mathcal{I}_n.\end{aligned} \tag{14}$$

The bias components are

$$B_{k0} = B_{k0}^{(1)} + B_{k0}^{(2)} \tag{15}$$

with

$$\begin{aligned}B_{k0}^{(1)} &:= \lim_{n\to\infty} \tfrac{1}{2\sqrt{N}} \mathrm{Tr}[\mathbf{J}_{11,0}^{-1}\mathbf{V}_{11,0}(\mathbf{J}_{11,0}^{-1})^\top \mathbf{G}_{k,0}],\\ B_{k0}^{(2)} &:= \lim_{n\to\infty} \tfrac{1}{2\sqrt{N}} \mathrm{Tr}[\mathbf{J}_{11,0}^{-1}\mathbf{V}_{11,0}(\mathbf{J}_{11,0}^{-1})^\top \mathbf{R}_{k,0}],\end{aligned}$$

and $B_0 = (B_{10},\ldots,B_{K0})^\top$. The sandwich-form asymptotic variance of $\widehat{\beta}$ is

$$\Omega_0 := \lim_{n\to\infty} N^{-1}\mathrm{J}_0^{-1}\begin{bmatrix}\mathrm{V}_{22,0} + \mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\mathbf{V}_{11,0}(\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1})^\top\\ -\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\mathbf{V}_{12,0} - (\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\mathbf{V}_{12,0})^\top\end{bmatrix}(\mathrm{J}_0^{-1})^\top. \tag{16}$$

Bias and variance of $\widehat{\delta}$. Define

$$\Delta_\beta(\boldsymbol{\alpha},\beta) := \frac{1}{N}\sum_{i=1}^n\sum_{j>i}\frac{\partial\Delta_{ij}}{\partial\beta}(\alpha_i,\alpha_j,\beta),\quad \Delta_{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\beta) := \frac{1}{N}\Big(\sum_{j\neq 1}\frac{\partial\Delta_{1j}}{\partial\alpha_1},\ldots,\sum_{j\neq n}\frac{\partial\Delta_{nj}}{\partial\alpha_n}\Big)^\top,$$

and the second-derivative matrices $\mathbf{R}_k^\mu$ and $\mathbf{G}_k^\mu$ analogously to $\mathbf{R}_k$ and $\mathbf{G}_k$ but with $\Delta_{ij,k}$ replacing $p_{ij}x_{ij,k}$. The bias terms are

$$\begin{aligned}B_{\alpha,k} &:= \lim_{n\to\infty}\frac{1}{2\sqrt{N}}\mathrm{Tr}\left[\mathbf{J}_{11,0}^{-1}\mathbf{V}_{11,0}\left(\mathbf{J}_{11,0}^{-1}\right)^\top(\mathbf{R}_{k,0}^\mu + \mathbf{G}_{k,0}^\mu)\right],\\ B_\beta &:= \lim_{n\to\infty}(\Delta_{\beta,0}^\top - \Delta_{\boldsymbol{\alpha},0}^\top\mathbf{J}_{11,0}^{-1}\mathbf{J}_{12,0})\mathrm{J}_0^{-1}B_0,\end{aligned} \tag{17}$$

and Σ_Δ is the asymptotic variance of $\sqrt{N}(\widehat{\delta} - \bar{\Delta}_n)$, given by

$$\Sigma_\Delta = \lim_{n\to\infty} \frac{1}{N} \left\{ \begin{aligned} &\mathbf{A}_0 \mathrm{V}_{22,0} \mathbf{A}_0^\top + \mathbf{C}_0 \mathbf{J}_{11,0}^{-1} \mathbf{V}_{11,0} (\mathbf{C}_0 \mathbf{J}_{11,0}^{-1})^\top \\ &- \mathbf{C}_0 \mathbf{J}_{11,0}^{-1} \mathbf{V}_{12,0} \mathbf{A}_0^\top - (\mathbf{C}_0 \mathbf{J}_{11,0}^{-1} \mathbf{V}_{12,0} \mathbf{A}_0^\top)^\top \end{aligned} \right\}, \tag{18}$$

where $\mathbf{A}_0 \coloneqq (\Delta_{\beta,0}^\top - \Delta_{\boldsymbol{\alpha},0}^\top \mathbf{J}_{11,0}^{-1} \mathbf{J}_{12,0}) \mathrm{J}_0^{-1}$ and $\mathbf{C}_0 \coloneqq \mathbf{A}_0 \mathbf{J}_{21,0} - N \Delta_{\boldsymbol{\alpha},0}^\top$.

B Supporting Lemmas

We state auxiliary results used in the proofs below. Their proofs are collected in the Supplemental Material.

Lemma 2 (Inverse approximation (Yan, 2019)). *Suppose an $n \times n$ matrix $\mathbf{A} = (a_{ij})_{n\times n}$ is invertible with all entries positive and $a_{ii} \geq \sum_{j\neq i} a_{ji}$. Let $\mathbf{B} = [\mathrm{diag}(a_{11}, \ldots, a_{nn})]^{-1}$, $\Delta_i = a_{ii} - \sum_{j\neq i} a_{ji}$, $M = \max\{\max_{i\neq j} a_{ij}, \max_i \Delta_i\}$, and $m = \min_{i\neq j} a_{ij}$. If $M \asymp 1$ and $m \asymp 1$, then $\|\mathbf{A}^{-1} - \mathbf{B}\|_{\max} = O(n^{-2})$.*

Lemma 3 (Diagonal approximation for $\mathbf{I}_{11}^{-1}$). *If Assumptions 1–6 hold, then uniformly over $(\boldsymbol{\alpha}, \beta) \in \mathbb{A} \times \mathbb{B}$, $\mathbf{D}_{ii}(\boldsymbol{\alpha}, \beta) \asymp n$, $\|\mathbf{I}_{11}^{-1}(\boldsymbol{\alpha}, \beta)\|_\infty = O(n^{-1})$, and $\|\mathbf{I}_{11}^{-1}(\boldsymbol{\alpha}, \beta) - \mathbf{D}^{-1}(\boldsymbol{\alpha}, \beta)\|_{\max} = O(n^{-2})$.*

Lemma 4 (Profiling weight bounds). *If Assumptions 1–6 hold, then uniformly over $(\boldsymbol{\alpha}, \beta) \in \mathbb{A} \times \mathbb{B}$ and all k, $\sup_i |w_{ki}| = O(1)$, $\sup_i \|\partial w_{ki}/\partial \beta\|_2 = O(1)$, $\sup_i |\partial w_{ki}/\partial \alpha_i| = O(1)$, and $\sup_{i\neq j} |\partial w_{ki}/\partial \alpha_j| = O(n^{-1})$.*

Lemma 5 (Well-definedness of limiting objects). *If Assumptions 1–5 hold, then the limits J_0, Ω_0, and B_0 exist as finite quantities, J_0 is invertible, and Ω_0 is positive definite. If Assumptions 1–6 hold, then the analogous limits I_0 and b_0 satisfy the same properties, with I_0 positive definite.*

We say an $n \times n$ matrix $\mathbf{A}$ belongs to the class $\mathcal{G}_n(\delta)$ if $\|\mathbf{A}\|_1 \leq 1$ and, for each $1 \leq i \neq j \leq n$, $\mathbf{A}_{ii} \geq \delta$ and $\mathbf{A}_{ij} \leq -\delta/(n-1)$.

Lemma 6 (Matrix contraction). *If $\mathbf{A}, \mathbf{B} \in \mathcal{G}_n(\delta)$, then*

$$\|\mathbf{AB}\|_1 \leq 1 - 2(n-2)(n-1)^{-1}\delta^2.$$

Lemma 7 (Deviation bound). *If Assumptions 1, 2, and 4 hold, then for any bounded array (λ_{ij}), $\Pr\left\{\max_i (n-1)^{-1} \left|\sum_{j\neq i} \lambda_{ij}(y_{ij} - p_{ij,0})\right| > C_1 \sqrt{6 \log n/(n-1)}\right\} \leq 2n^{-2}$.*

Define $\boldsymbol{\eta}_0 = \sum_{k=1}^{n}(\widehat{\alpha}_k(\beta_0) - \alpha_{k0})\frac{\partial \mathbf{J}_{11}(\boldsymbol{\alpha}_0,\beta_0)}{\partial \alpha_k}[\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0]$.

Lemma 8 (Higher-order expansion of $\widehat{\boldsymbol{\alpha}}$)**.** *If Assumptions 1, 2, and 4 hold, then:*

(a) $\|\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0\|_\infty = O_p\left(\sqrt{\log n/n}\right)$.

(b) $\mathrm{plim}_{n\to\infty}\frac{1}{2\sqrt{N}}\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0 = B_0^{(1)}$.

(c) $\left\|\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0 + \mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0} + \frac{1}{2}\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0\right\|_\infty = O_p\left((\log n)^{3/2}/n^{3/2}\right)$.

Define $S_n(\beta) \coloneqq N^{-1}m_2(\widehat{\boldsymbol{\alpha}}(\beta),\beta)$. Recall $\bar{S}_n(\beta)$ from (6).

Lemma 9 (Uniform convergence)**.** *If Assumptions 1–5 hold, then*

$$\sup_{\beta\in\mathbb{B}}\left\|S_n(\beta) - \bar{S}_n(\beta)\right\|_2 \xrightarrow{p} 0.$$

Lemma 10 (One-step score bounds)**.** *If Assumptions 1–6 hold, then:*

(a) $N^{-1/2}\nabla_{\boldsymbol{\alpha}^\top}s_n(\widehat{\boldsymbol{\alpha}}(\beta_0),\beta_0)[\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0] \xrightarrow{p} b_0$.

(b) $N^{-1}\nabla_{\beta^\top}s_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}),\bar{\beta}) + N^{-1}\nabla_{\boldsymbol{\alpha}^\top}s_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}),\bar{\beta})\partial\widehat{\boldsymbol{\alpha}}(\bar{\beta})/\partial\beta^\top + \mathrm{I}_0 \xrightarrow{p} 0$,

for any $\bar{\beta}$ between β_0 and its $\sqrt{N}$-consistent estimator.

C Proofs

This appendix contains the proofs of Lemma 1 and Theorems 1–4. Proofs of the supporting lemmas stated in Appendix B are deferred to Section A of the Supplemental Material.

C.1 Proof of Lemma 1

Proof of Lemma 1. First, suppose a solution to $\mathbf{m}_1(\boldsymbol{\alpha},\beta) = 0$ exists. Let $\mathbf{G}^\circ(\boldsymbol{\alpha},\widehat{\boldsymbol{\alpha}})$ be the matrix whose (i,j)th element is

$$[\mathbf{G}^\circ(\boldsymbol{\alpha},\widehat{\boldsymbol{\alpha}})]_{ij} = \int_0^1 \frac{\partial r_i}{\partial \alpha_j}(t\boldsymbol{\alpha} + (1-t)\widehat{\boldsymbol{\alpha}})dt.$$

Then, by an integral form of the mean value theorem, we have

$$\mathbf{r}(\boldsymbol{\alpha}) - \mathbf{r}(\widehat{\boldsymbol{\alpha}}) = \mathbf{G}^\circ(\boldsymbol{\alpha},\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha} - \widehat{\boldsymbol{\alpha}}).$$

Notice that for $i \neq j$, $\partial r_j/\partial \alpha_i = -(n-1)^{-1} p_{ji}^{(0,1,0)}(\boldsymbol{\alpha}, \beta) < 0$; while for each i, $\partial r_i/\partial \alpha_i = 1 - (n-1)^{-1} \sum_{j \neq i} p_{ij}^{(1,0,0)}(\boldsymbol{\alpha}, \beta) > 0$. Moreover, for each i,

$$\sum_{j=1}^{n} \left| \frac{\partial r_j}{\partial \alpha_i} \right| = \frac{\partial r_i}{\partial \alpha_i} - \sum_{j \neq i} \frac{\partial r_j}{\partial \alpha_i} \equiv 1.$$

For each i and any $\boldsymbol{\alpha}$, this proves $\sum_{j=1}^{n} |[\mathbf{G}^{\circ}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}})]_{ji}| = 1$, i.e., $\|\mathbf{G}^{\circ}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}})\|_1 = 1$. By Assumptions 2 and 4, the derivatives are uniformly bounded and $\partial r_j/\partial \alpha_i < 0$ for $i \neq j$, while $\partial r_i/\partial \alpha_i > 0$. By Assumption 4, there exists a constant $\delta \in (0,1)$ such that $[\mathbf{G}^{\circ}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}})]_{ii} \geq \delta$ and $[\mathbf{G}^{\circ}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}})]_{ij} \leq -\delta/(n-1)$ for all $i \neq j$. Therefore, we have $\mathbf{G}^{\circ}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}) \in \mathcal{G}_n(\delta)$.

By the updating algorithm (2), after every two updates, we have

$$\begin{aligned}
\|\boldsymbol{\alpha}^{k+2}(\beta) - \widehat{\boldsymbol{\alpha}}(\beta)\|_1 &= \|\mathbf{r}(\mathbf{r}(\boldsymbol{\alpha}^{k}(\beta))) - \mathbf{r}(\mathbf{r}(\widehat{\boldsymbol{\alpha}}(\beta)))\|_1 \\
&= \|\mathbf{G}^{\circ}(\mathbf{r}(\boldsymbol{\alpha}^{k}(\beta)), \widehat{\boldsymbol{\alpha}}(\beta))(\mathbf{r}(\boldsymbol{\alpha}^{k}(\beta)) - \widehat{\boldsymbol{\alpha}}(\beta))\|_1 \\
&= \|\mathbf{G}^{\circ}(\mathbf{r}(\boldsymbol{\alpha}^{k}(\beta)), \widehat{\boldsymbol{\alpha}}(\beta))\mathbf{G}^{\circ}(\boldsymbol{\alpha}^{k}(\beta), \widehat{\boldsymbol{\alpha}}(\beta))(\boldsymbol{\alpha}^{k}(\beta) - \widehat{\boldsymbol{\alpha}}(\beta))\|_1 \\
&\leq \|\mathbf{G}^{\circ}(\mathbf{r}(\boldsymbol{\alpha}^{k}(\beta)), \widehat{\boldsymbol{\alpha}}(\beta))\mathbf{G}^{\circ}(\boldsymbol{\alpha}^{k}(\beta), \widehat{\boldsymbol{\alpha}}(\beta))\|_1 \|\boldsymbol{\alpha}^{k}(\beta) - \widehat{\boldsymbol{\alpha}}(\beta)\|_1 \\
&\leq \left(1 - \frac{2(n-2)}{n-1}\delta^2\right) \|\boldsymbol{\alpha}^{k}(\beta) - \widehat{\boldsymbol{\alpha}}(\beta)\|_1,
\end{aligned}$$

where the first equality holds by the fact that $\widehat{\boldsymbol{\alpha}}(\beta) = \mathbf{r}(\widehat{\boldsymbol{\alpha}}(\beta))$ which implies $\widehat{\boldsymbol{\alpha}}(\beta)$ is the fixed point of the updating function, and the last inequality holds by Lemma 6. We write $\bar{\delta} \coloneqq 1 - \frac{2(n-2)}{n-1}\delta^2$, and the second inequality of Lemma 1 follows.

By this result, $\mathbf{r}(\boldsymbol{\alpha})$ is a contraction mapping for $(\boldsymbol{\alpha}, \beta) \in \mathbb{A} \times \mathbb{B}$. So, if there exists a solution $\widehat{\boldsymbol{\alpha}}(\beta) \in \mathbb{A}$, the solution is unique. Now we show the existence of the solution, where the main technique is adapted from Yan, Qin, and Wang (2016) and Yan, Jiang, Fienberg, and Leng (2019). Define a sequence of Newton iterations $\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} - \mathbf{J}_{11}^{-1}(\boldsymbol{\alpha}^{(k)}, \beta)\mathbf{m}_1(\boldsymbol{\alpha}^{(k)}, \beta)$, and choose the initial value as $\boldsymbol{\alpha}^{(0)} = \boldsymbol{\alpha}_0$. Following Proposition A.1 of Yan, Qin, and Wang (2016), in a convex subset $\mathbb{D} \subset \mathbb{A}$ that contains $\boldsymbol{\alpha}_0$, to obtain the existence of the solution it is sufficient to establish three facts: (1) $\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)$ is Lipschitz continuous with Lipschitz constant of order $O(n)$, (2) $\left\|\mathbf{J}_{11}^{-1}(\boldsymbol{\alpha}_0, \beta)\right\|_\infty = O(n^{-1})$, and (3) $\left\|\mathbf{J}_{11}^{-1}(\boldsymbol{\alpha}_0, \beta)\mathbf{m}_1(\boldsymbol{\alpha}_0, \beta)\right\|_\infty = O(\|\beta - \beta_0\|_2)$.

For the first fact, we calculate the derivative of $\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)$ with respect to $\boldsymbol{\alpha}$:

$$
\begin{aligned}
\frac{\partial \mathbf{J}_{11,ij}}{\partial \alpha_k} =& - \mathbf{1}\{i = j = k\} \textstyle\sum_{l \neq i} p_{il}^{(2,0,0)} - \mathbf{1}\{i = j \neq k\} p_{ik}^{(1,1,0)} \\
& - \mathbf{1}\{j \neq i = k\} p_{ij}^{(1,1,0)} - \mathbf{1}\{i \neq j = k\} p_{ij}^{(0,2,0)},
\end{aligned}
$$

which implies that $\max_i \sum_{j,k} \left| \int_0^1 \frac{\partial \mathbf{J}_{11,ij}(t\boldsymbol{\alpha}_1 + (1-t)\boldsymbol{\alpha}_2)}{\partial \alpha_k} dt \right| = O(n)$. Hence $\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)$ is Lipschitz continuous with Lipschitz constant $O(n)$. The second fact is a direct application of the inverse approximation Lemma 2. Finally, the third result follows from

$$
\begin{aligned}
& \left\| [\mathbf{J}_{11}(\boldsymbol{\alpha}_0, \beta)]^{-1} \mathbf{m}_1(\boldsymbol{\alpha}_0, \beta) \right\|_\infty \\
\leq \ & \left\| [\mathbf{J}_{11}(\boldsymbol{\alpha}_0, \beta)]^{-1} \mathbf{m}_{1,0} \right\|_\infty + \left\| [\mathbf{J}_{11}(\boldsymbol{\alpha}_0, \beta)]^{-1} \left[\mathbf{m}_1(\boldsymbol{\alpha}_0, \beta) - \mathbf{m}_{1,0} \right] \right\|_\infty \\
\leq \ & o_p(1) + O\left(\|\beta - \beta_0\|_2 \right),
\end{aligned}
$$

where the first inequality holds by the triangle inequality and the second inequality is true by Lemma 7 and the Lipschitz continuity of $p(\cdot)$ under Assumption 4. In particular, for any β in a sufficiently small neighborhood of β_0, the right-hand side is $o_p(1)$. Then, by an application of Proposition A.1 of Yan, Qin, and Wang (2016), we have $\lim_{k \to \infty} \boldsymbol{\alpha}^{(k)}$ exists and the limit equals $\widehat{\boldsymbol{\alpha}}(\beta)$ if $\|\beta - \beta_0\|_2 < c$ for some constant $c > 0$. $\square$

C.2 Proof of Theorem 1

Proof of Theorem 1. *Consistency.* Recall $S_n(\beta)$ and $\bar{S}_n(\beta)$ from Lemma 9. By Assumption 5, $\widehat{\beta}$ and β_0 are unique solutions to $S_n(\beta) = 0$ and $\bar{S}_n(\beta) = 0$, respectively. By Lemma 9,

$$
\left\| \bar{S}_n(\widehat{\beta}) \right\|_2 = \left\| \bar{S}_n(\widehat{\beta}) - S_n(\widehat{\beta}) \right\|_2 \leq \sup_{\beta \in \mathbb{B}} \left\| S_n(\beta) - \bar{S}_n(\beta) \right\|_2 \xrightarrow{p} 0. \tag{19}
$$

Fix $\delta > 0$. By Assumption 5, there exists an $\epsilon > 0$ such that $\|\beta - \beta_0\|_2 \geq \delta$ implies $\left\| \bar{S}_n(\beta) \right\|_2 \geq \epsilon$, hence

$$
\Pr\left(\left\| \widehat{\beta} - \beta_0 \right\|_2 \geq \delta \right) \leq \Pr\left(\left\| \bar{S}_n(\widehat{\beta}) \right\|_2 \geq \epsilon \right) \leq \Pr\left(\sup_{\beta \in \mathbb{B}} \left\| S_n(\beta) - \bar{S}_n(\beta) \right\|_2 \geq \epsilon \right) \to 0
$$

by (19).

We turn to the proof of the convergence of $\widehat{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}_0$ in the ℓ_∞ norm. By the integral type mean-value theorem, we have

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0 &= -\left[\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}},\boldsymbol{\alpha}_0)\right]^{-1}\mathbf{m}_1(\boldsymbol{\alpha}_0,\widehat{\beta})\\ &= -\left[\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}},\boldsymbol{\alpha}_0)\right]^{-1}\mathbf{m}_{1,0}-\left[\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}},\boldsymbol{\alpha}_0)\right]^{-1}\left(\mathbf{m}_1(\boldsymbol{\alpha}_0,\widehat{\beta})-\mathbf{m}_{1,0}\right)\end{aligned}$$

Following the proof of Lemma 8, we have $\|[\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}},\boldsymbol{\alpha}_0)]^{-1}\|_\infty = O(n^{-1})$ and $\|\mathbf{m}_{1,0}\|_\infty = O_p(\sqrt{n\log n})$, hence $\left\|[\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}},\boldsymbol{\alpha}_0)]^{-1}\mathbf{m}_{1,0}\right\|_\infty \xrightarrow{p} 0$. Thus, we only need to show that $O(n^{-1})\cdot\|\mathbf{m}_1(\boldsymbol{\alpha}_0,\widehat{\beta})-\mathbf{m}_{1,0}\|_\infty \xrightarrow{p} 0$. Notice that

$$\begin{aligned}\|\mathbf{m}_1(\boldsymbol{\alpha}_0,\widehat{\beta})-\mathbf{m}_{1,0}\|_\infty &= \max_{1\le i\le n}\left|\sum\nolimits_{j\neq i}\left[p_{ij}(\boldsymbol{\alpha}_0,\widehat{\beta})-p_{ij,0}\right]\right|\\ &\le \max_{1\le i\le n}\left\|\sum\nolimits_{j\neq i}p_{ij}^{(0,0,1)}(\boldsymbol{\alpha}_0,\bar{\beta})x_{ij}\right\|_2\cdot\|\widehat{\beta}-\beta_0\|_2 = o_p(n),\end{aligned}$$

where we use a Taylor expansion of $p_{ij}(\boldsymbol{\alpha}_0,\beta)$ around β_0 ($\bar{\beta}$ is the mean value which may vary with i) and the fact that $p_{ij}^{(0,0,1)}$ is bounded by Assumption 4.

Asymptotic normality. By Lemma 5, $N^{-1}\mathrm{J}_{n,0} \xrightarrow{p} \mathrm{J}_0$, a finite positive-definite matrix. Combined with $\widehat{\beta} \xrightarrow{p} \beta_0$, $\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|_\infty = o_p(1)$ from the consistency part, and the continuity of $N^{-1}\mathrm{J}_n$ in $(\boldsymbol{\alpha},\beta)$, we have $N^{-1}\mathrm{J}_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}),\bar{\beta}) \xrightarrow{p} \mathrm{J}_0$ for any $\bar{\beta}$ between $\widehat{\beta}$ and β_0.

By a first-order Taylor expansion of $m_n(\widehat{\beta}) = m_2(\widehat{\boldsymbol{\alpha}}(\widehat{\beta}),\widehat{\beta})$ around β_0, we have

$$m_n(\widehat{\beta})-m_n(\beta_0) = \mathrm{J}_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}),\bar{\beta})(\widehat{\beta}-\beta_0),$$

where $\bar{\beta}$ is the mean-value between $\widehat{\beta}$ and β_0. By $m_n(\widehat{\beta})=0$, we obtain

$$\begin{aligned}\sqrt{N}(\widehat{\beta}-\beta_0) &= -\left[N^{-1}\mathrm{J}_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}),\bar{\beta})\right]^{-1}\frac{1}{\sqrt{N}}m_2(\widehat{\boldsymbol{\alpha}}(\beta_0),\beta_0)\\ &= -\mathrm{J}_0^{-1}\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{j>i}\left[y_{ij}-p_{ij}(\widehat{\boldsymbol{\alpha}}(\beta_0),\beta_0)\right]x_{ij}\right\}+o_p(1) \qquad (20)\end{aligned}$$

by the definition of J_0. The bracketed term depends on $\widehat{\boldsymbol{\alpha}}(\beta_0)$ and requires expansion before a CLT applies. Setting $\boldsymbol{\zeta} = \widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0$, a third-order Taylor expansion of $m_2(\widehat{\boldsymbol{\alpha}}(\beta_0),\beta_0)$ in $\boldsymbol{\alpha}$ around $\boldsymbol{\alpha}_0$ gives

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{j>i}\left[y_{ij}-p_{ij}(\widehat{\boldsymbol{\alpha}}(\beta_0),\beta_0)\right]x_{ij}\\ &= \frac{1}{\sqrt{N}}m_{2,0}+\frac{1}{\sqrt{N}}\mathbf{J}_{21,0}\boldsymbol{\zeta}-\frac{1}{2\sqrt{N}}\sum_{k=1}^{n}\zeta_k\sum_{i=1}^{n}\sum_{j>i}\frac{\partial^2 p_{ij}(\boldsymbol{\alpha}_0,\beta_0)}{\partial\alpha_k\partial\boldsymbol{\alpha}^\top}\boldsymbol{\zeta}\,x_{ij}\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{6\sqrt{N}} \sum_{k=1}^{n}\sum_{l=1}^{n} \zeta_k \zeta_l \sum_{i=1}^{n}\sum_{j>i} \frac{\partial^3 p_{ij}(\bar{\boldsymbol{\alpha}}, \beta_0)}{\partial \alpha_k \partial \alpha_l \partial \boldsymbol{\alpha}^\top} \boldsymbol{\zeta}\, x_{ij} \\
&=: (I) + (II) + (III) + (IV)\,. \qquad (21)
\end{aligned}
$$

Since p_{ij} depends only on α_i and α_j, and $\sup_i |\widehat{\alpha}_i(\beta_0) - \alpha_{i0}| = O_p(\sqrt{(\log n)/n})$ by Lemma 8, the third-order remainder satisfies $(IV) = O_p(N^{-1/2} \cdot N \cdot (\log n/n)^{3/2}) = o_p(1)$ under Assumptions 2 and 4.

For (III), substituting the asymptotic linear approximation of $\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0$, its kth entry equals $-\frac{1}{2\sqrt{N}}\mathrm{Tr}[\mathbf{J}_{11,0}^{-1}\mathbf{V}_{11,0}(\mathbf{J}_{11,0}^{-1})^\top \mathbf{R}_{k,0}] + o_p(1)$, where $\mathbf{R}_k$ for $k = 1, \ldots, K$ is defined in (14). Recalling the definition of $B_{k0}^{(2)}$ in (15), it follows that $(III) = -B_0^{(2)} + o_p(1)$, where $B_0^{(2)} := (B_{10}^{(2)}, \ldots, B_{K0}^{(2)})^\top$.

We directly combine the rest of terms, (I) and (II), to obtain

$$
\begin{aligned}
&\sqrt{N}(\widehat{\beta} - \beta_0) - \mathrm{J}_0^{-1} B_0 \\
&= -\,\mathrm{J}_0^{-1} \left\{ \frac{1}{\sqrt{N}} m_{2,0} + \frac{1}{\sqrt{N}} \mathbf{J}_{21,0} [\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0] \right\} + \mathrm{J}_0^{-1} B_0^{(2)} - \mathrm{J}_0^{-1} B_0 + o_p(1) \\
&= -\,\mathrm{J}_0^{-1} \left\{ \frac{1}{\sqrt{N}} m_{2,0} - \frac{1}{\sqrt{N}} \mathbf{J}_{21,0} \mathbf{J}_{11,0}^{-1} \mathbf{m}_{1,0} \right\} + o_p(1), \qquad (22)
\end{aligned}
$$

where we use the facts that $B_0^{(1)} := (B_{10}^{(1)}, \ldots, B_{K0}^{(1)})^\top$, $\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0 = -\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0} - \frac{1}{2}\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0 - \frac{1}{6}\mathbf{J}_{11,0}^{-1}\boldsymbol{\rho}_0$ by Lemma 8(c), $\frac{1}{2\sqrt{N}}\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0 - B_0^{(1)} = o_p(1)$, and $\|\mathbf{J}_{11,0}^{-1}\boldsymbol{\rho}_0\|_\infty = O((\log n)^{3/2}/n^{3/2})$ by part (c) of Lemma 8. To apply the CLT, we verify the Lindeberg condition. Define

$$
\begin{aligned}
\frac{1}{\sqrt{N}} \sum_{i=1}^{n}\sum_{j>i} \xi_{ij} &:= -\,\mathrm{J}_0^{-1} \frac{1}{\sqrt{N}} \left\{ \sum_{i=1}^{n}\sum_{j>i} (y_{ij} - p_{ij,0}) x_{ij} - \mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0} \right\} \\
&= -\,\mathrm{J}_0^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{n}\sum_{j>i} (y_{ij} - p_{ij,0}) \widetilde{x}_{ij}, \qquad (23)
\end{aligned}
$$

where $\widetilde{x}_{ij}$ collects the two multipliers of $(y_{ij} - p_{ij,0})$ from the definitions in Appendix A and is uniformly bounded by Assumptions 2 and 4. Since the centered Bernoulli $y_{ij} - p_{ij,0}$ are independent across dyads and bounded in $[-1, 1]$, the Lindeberg condition holds, yielding $\sqrt{N}(\widehat{\beta} - \beta_0) - \mathrm{J}_0^{-1} B_0 \xrightarrow{d} \mathcal{N}(0, \Omega_0)$, where Ω_0 is defined in (16). □

C.3 Proof of Theorem 2

Proof of Theorem 2. By definition, $\widehat{\beta}_{\mathrm{OS}} = \widehat{\beta} + \mathrm{I}_n(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})^{-1} s_n(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})$ with the JMM estimator $(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})$. A first-order Taylor expansion of $s_n(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})$ around β_0, followed by a first-order expansion of $s_n(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0)$ around $\boldsymbol{\alpha}_0$, together with $N^{-1}\mathrm{I}_n(\widehat{\boldsymbol{\alpha}}, \widehat{\beta}) \xrightarrow{p} \mathrm{I}_0$, yields

$$
\begin{aligned}
&\sqrt{N}(\widehat{\beta}_{\mathrm{OS}} - \beta_0) \\
=& \frac{1}{\sqrt{N}} \mathrm{I}_0^{-1} s_n(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0) \\
&+ \mathrm{I}_0^{-1} \left[\frac{1}{N} \nabla_{\beta^\top} s_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta}) + \frac{1}{N} \nabla_{\boldsymbol{\alpha}^\top} s_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta}) \frac{\partial \widehat{\boldsymbol{\alpha}}(\bar{\beta})}{\partial \beta^\top} + \mathrm{I}_0 \right] \sqrt{N}(\widehat{\beta} - \beta_0) + o_p(1) \\
=& \frac{1}{\sqrt{N}} \mathrm{I}_0^{-1} s_{n,0} + \frac{1}{\sqrt{N}} \mathrm{I}_0^{-1} \nabla_{\boldsymbol{\alpha}^\top} s_n(\bar{\boldsymbol{\alpha}}, \beta_0)(\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0) \\
&+ \mathrm{I}_0^{-1} \left[\frac{1}{N} \nabla_{\beta^\top} s_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta}) + \frac{1}{N} \nabla_{\boldsymbol{\alpha}^\top} s_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta}) \frac{\partial \widehat{\boldsymbol{\alpha}}(\bar{\beta})}{\partial \beta^\top} + \mathrm{I}_0 \right] \sqrt{N}(\widehat{\beta} - \beta_0) + o_p(1).
\end{aligned}
\tag{24}
$$

By Lemma 10, we have $\frac{1}{\sqrt{N}} \nabla_{\boldsymbol{\alpha}^\top} s_n(\bar{\boldsymbol{\alpha}}, \beta_0)(\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0) \xrightarrow{p} b_0$ and

$$
\frac{1}{N} \nabla_{\beta^\top} s_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta}) + \frac{1}{N} \nabla_{\boldsymbol{\alpha}^\top} s_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta}) \frac{\partial \widehat{\boldsymbol{\alpha}}(\bar{\beta})}{\partial \beta^\top} + \mathrm{I}_0 = o_p(1).
$$

Hence, using the result that $\sqrt{N}(\widehat{\beta} - \beta_0) = O_p(1)$ by Theorem 1, we simplify (24) as

$$
\begin{aligned}
\sqrt{N}(\widehat{\beta}_{\mathrm{OS}} - \beta_0) &= \frac{1}{\sqrt{N}} \mathrm{I}_0^{-1} s_{n,0} + \mathrm{I}_0^{-1} b_0 + o_p(1) \\
&= \frac{1}{\sqrt{N}} \mathrm{I}_0^{-1} \sum_{i=1}^{n} \sum_{j>i} s_{ij,0} + \mathrm{I}_0^{-1} b_0 + o_p(1),
\end{aligned}
\tag{25}
$$

where $s_{ij,0}$ is dyad (i,j)'s contribution to the asymptotic representation. Then, by the Lindeberg-Feller CLT, as in the proof of Theorem 1, we have the stated asymptotic normality. □

C.4 Proof of Theorem 3

Proof. We define $\widehat{\beta}_{T_n} = \frac{1}{T_n} \sum_{s=1}^{T_n} \widehat{\beta}_{\mathrm{OS-SJ}}^{(s)}$ as the average of all possible OS-SJ estimators. Let $\mathcal{F}_n$ be the σ-algebra generated by all observed information. It is clear that $\widehat{\beta}_{T_n}$ is $\mathcal{F}_n$-measurable. Let $\mathbb{E}^*$ represent the expectation over randomness from the random splits conditional on $\mathcal{F}_n$. Our proof contains two immediate results: (i)

$\sqrt{N}(\widehat{\beta}_{T_n} - \beta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{I}_0^{-1})$ as $n \to \infty$; (ii) $\sqrt{N}(\widehat{\beta}_{\mathrm{BG}} - \widehat{\beta}_{T_n}) \xrightarrow{p} 0$ as $n \to \infty$ and $\widetilde{T}_n \to \infty$. Theorem 3 follows by a combination of these two results.

Step (i). By (25), we have

$$\sqrt{N}(\widehat{\beta}_{\mathrm{OS}} - \beta_0) = \mathrm{I}_0^{-1} b_0 + \frac{1}{\sqrt{N}} \mathrm{I}_0^{-1} \sum_{(i,j) \in \mathcal{I}_n \times \mathcal{I}_n; j>i} s_{ij,0} + \mathcal{R}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0),$$

where $\mathcal{R}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0)$ is a residual term of order $o_p(1)$, as shown in the proof of Lemma 10. Thus, the one-step estimators based on sub-networks are (using $N/4$ to approximate the sub-network dyad count $\binom{n/2}{2}$, with the $O(n^{-1})$ discrepancy absorbed into $\mathcal{R}$)

$$\sqrt{N/4}(\widehat{\beta}_{\mathrm{OS},1}^{(t)} - \beta_0) = \mathrm{I}_0^{-1} b_0 + \frac{1}{\sqrt{N/4}} \mathrm{I}_0^{-1} \sum_{(i,j) \in \mathcal{I}_{1,n}^{(t)} \times \mathcal{I}_{1,n}^{(t)}; j>i} s_{ij,0} + \mathcal{R}(\mathbf{y}_1^{(t)}, \mathbf{x}_1^{(t)}, \boldsymbol{\alpha}_{0,1}^{(t)}),$$

$$\sqrt{N/4}(\widehat{\beta}_{\mathrm{OS},2}^{(t)} - \beta_0) = \mathrm{I}_0^{-1} b_0 + \frac{1}{\sqrt{N/4}} \mathrm{I}_0^{-1} \sum_{(i,j) \in \mathcal{I}_{2,n}^{(t)} \times \mathcal{I}_{2,n}^{(t)}; j>i} s_{ij,0} + \mathcal{R}(\mathbf{y}_2^{(t)}, \mathbf{x}_2^{(t)}, \boldsymbol{\alpha}_{0,2}^{(t)}),$$

where $\boldsymbol{\alpha}_{0,1}^{(t)}$ is the sub-vector of $\boldsymbol{\alpha}_0$ indexed by $\mathcal{I}_{1,n}^{(t)}$ and similarly for $\boldsymbol{\alpha}_{0,2}^{(t)}$. Hence, we have

$$\sqrt{N}(\widehat{\beta}_{\mathrm{OS-SJ}}^{(t)} - \beta_0) = \frac{2}{\sqrt{N}} \mathrm{I}_0^{-1} \sum_{(i,j) \in \mathcal{I}_{1,n}^{(t)} \times \mathcal{I}_{2,n}^{(t)}} s_{ij,0} + \mathcal{R}^{(t)}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0), \tag{26}$$

with $\mathcal{R}^{(t)}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0) := 2\mathcal{R}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0) - \left[\mathcal{R}(\mathbf{y}_1^{(t)}, \mathbf{x}_1^{(t)}, \boldsymbol{\alpha}_{0,1}^{(t)}) + \mathcal{R}(\mathbf{y}_2^{(t)}, \mathbf{x}_2^{(t)}, \boldsymbol{\alpha}_{0,2}^{(t)})\right]$, which is also $o_p(1)$ for each fixed t. Because the T_n equal splits are averaged uniformly,

$$\frac{1}{T_n} \sum_{t=1}^{T_n} \mathcal{R}^{(t)}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0) = \mathbb{E}^*\left[\mathcal{R}^{(1)}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0) \mid \mathcal{F}_n\right].$$

By Jensen's inequality,

$$\begin{aligned} \left\| \frac{1}{T_n} \sum_{t=1}^{T_n} \mathcal{R}^{(t)} \right\|_2 &\le 2 \left\| \mathcal{R}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0) \right\|_2 \\ &\quad + \mathbb{E}^*\left[\left\| \mathcal{R}(\mathbf{y}_1^{(1)}, \mathbf{x}_1^{(1)}, \boldsymbol{\alpha}_{0,1}^{(1)}) \right\|_2 + \left\| \mathcal{R}(\mathbf{y}_2^{(1)}, \mathbf{x}_2^{(1)}, \boldsymbol{\alpha}_{0,2}^{(1)}) \right\|_2 \mid \mathcal{F}_n \right]. \end{aligned}$$

The full-sample remainder is $o_p(1)$ by (25), and the same remainder calculations as in the proof of Lemma 10 apply to a generic half-sample because the primitive bounds there depend only on the uniform constants in the assumptions and on the subsample size, which is $n/2 \to \infty$. Therefore,

$$\frac{1}{T_n}\sum_{t=1}^{T_n} \mathcal{R}^{(t)}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0) = o_p(1).$$

Then, taking the average of (26) over all $1 \le t \le T_n$ yields

$$\begin{aligned}
\sqrt{N}(\widehat{\beta}_{T_n} - \beta_0) &= \frac{2}{\sqrt{N}}\mathrm{I}_0^{-1}\frac{1}{T_n}\sum_{t=1}^{T_n}\sum_{(i,j)\in\mathcal{I}_{1,n}^{(t)}\times\mathcal{I}_{2,n}^{(t)}} s_{ij,0} + \frac{1}{T_n}\sum_{t=1}^{T_n}\mathcal{R}^{(t)}(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}_0) \\
&= \frac{2}{\sqrt{N}}\mathrm{I}_0^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\frac{\binom{n-2}{n/2-1}}{\binom{n}{n/2}}s_{ij,0} + o_p(1) \\
&= \frac{1}{\sqrt{N}} \times \frac{n}{n-1}\mathrm{I}_0^{-1}\sum_{i=1}^{n}\sum_{j>i} s_{ij,0} + o_p(1) \xrightarrow{d} \mathcal{N}(0, \mathrm{I}_0^{-1}),
\end{aligned} \tag{27}$$

where the second equality holds because for each $(i, j), i \neq j$, there are $\binom{n-2}{n/2-1}$ different splits containing them. This proves the first result.

Step (ii). Conditional on $\mathcal{F}_n$, random draws $\widehat{\beta}^{(t)}_{\mathrm{OS-SJ}}$ for $t = 1, \ldots, \widetilde{T}_n$ are independent and uniformly distributed over $\{\widehat{\beta}^{(s)}_{\mathrm{OS-SJ}}\}_{s=1}^{T_n}$. Thus, $\mathbb{E}^*[\widehat{\beta}^{(t)}_{\mathrm{OS-SJ}}] = \widehat{\beta}_{T_n}$ and $\mathbb{E}^*[\|\widehat{\beta}^{(t)}_{\mathrm{OS-SJ}} - \widehat{\beta}_{T_n}\|_2^2] = \frac{1}{T_n}\sum_{s=1}^{T_n}\|\widehat{\beta}^{(s)}_{\mathrm{OS-SJ}} - \widehat{\beta}_{T_n}\|_2^2 =: \sigma_n^2$. Note that σ_n^2 is $\mathcal{F}_n$-measurable and $\sigma_n^2 = O_p(N^{-1})$ by (26) and (27). By Markov's inequality, for any $\epsilon > 0$

$$\Pr(\|\sqrt{N}(\widehat{\beta}_{\mathrm{BG}} - \widehat{\beta}_{T_n})\|_2 \ge \epsilon|\mathcal{F}_n) \le \frac{N\mathbb{E}^*[\|\widehat{\beta}_{\mathrm{BG}} - \widehat{\beta}_{T_n}\|_2^2]}{\epsilon^2} = \frac{N\sigma_n^2}{\widetilde{T}_n\epsilon^2}.$$

Hence, $\Pr(\|\sqrt{N}(\widehat{\beta}_{\mathrm{BG}} - \widehat{\beta}_{T_n})\|_2 \ge \epsilon) \le \mathbb{E}[\widetilde{T}_n^{-1}\epsilon^{-2}N\sigma_n^2] = O(\widetilde{T}_n^{-1}) = o(1)$ as $n \to \infty$ and $\widetilde{T}_n \to \infty$ for any $\epsilon > 0$. This proves the second result.

Combining Step (i) and Step (ii), we have

$$\sqrt{N}(\widehat{\beta}_{\mathrm{BG}} - \beta_0) = \sqrt{N}(\widehat{\beta}_{\mathrm{BG}} - \widehat{\beta}_{T_n}) + \sqrt{N}(\widehat{\beta}_{T_n} - \beta_0) = \sqrt{N}(\widehat{\beta}_{T_n} - \beta_0) + o_p(1) \xrightarrow{d} \mathcal{N}(0, \mathrm{I}_0^{-1})$$

as $\widetilde{T}_n \to \infty$ and $n \to \infty$. □

C.5 Proof of Theorem 4

Proof of Theorem 4. We decompose $\widehat{\delta} - \delta_0 = \left(\widehat{\delta} - \bar{\Delta}_n\right) + \left(\bar{\Delta}_n - \delta_0\right)$. The second term is a U-statistic:

$$\bar{\Delta}_n - \delta_0 = N^{-1}\sum_{i=1}^{n}\sum_{j>i}\left[\Delta_{ij}(\alpha_{i0}, \alpha_{j0}, \beta_0) - \mathbb{E}\Delta_{ij}(\alpha_{i0}, \alpha_{j0}, \beta_0)\right]$$

with kernel $\Delta_{ij}(\alpha_{i0}, \alpha_{j0}, \beta_0) - \mathbb{E}\Delta_{ij}(\alpha_{i0}, \alpha_{j0}, \beta_0)$. By Theorem 12.3 of van der Vaart (2000), we have

$$\sqrt{n}(\bar{\Delta}_n - \delta_0) \xrightarrow{d} \mathcal{N}(0, 4\Sigma_\delta), \tag{28}$$

where $\Sigma_\delta = \mathrm{Cov}(\Delta_{ij}(\alpha_{i0}, \alpha_{j0}, \beta_0), \Delta_{ik}(\alpha_{i0}, \alpha_{k0}, \beta_0))$. Next, for the first term, notice that $\widehat{\boldsymbol{\alpha}} \equiv \widehat{\boldsymbol{\alpha}}(\widehat{\beta})$ and we can decompose it as

$$\begin{aligned}
\sqrt{N}\left(\widehat{\delta} - \bar{\Delta}_n\right) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j>i} \left[\Delta_{ij}(\widehat{\alpha}_i(\widehat{\beta}), \widehat{\alpha}_j(\widehat{\beta}), \widehat{\beta}) - \Delta_{ij}(\alpha_{i0}, \alpha_{j0}, \beta_0)\right] \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j>i} \left[\Delta_{ij}(\widehat{\alpha}_i(\widehat{\beta}), \widehat{\alpha}_j(\widehat{\beta}), \widehat{\beta}) - \Delta_{ij}(\widehat{\alpha}_i(\beta_0), \widehat{\alpha}_j(\beta_0), \beta_0)\right] \\
&\quad + \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j>i} [\Delta_{ij}(\widehat{\alpha}_i(\beta_0), \widehat{\alpha}_j(\beta_0), \beta_0) - \Delta_{ij}(\alpha_{i0}, \alpha_{j0}, \beta_0)] \\
&=: U_1 + U_2,
\end{aligned}$$

where U_1 captures the variation from $\widehat{\beta}$ and U_2 captures the variation from $\widehat{\boldsymbol{\alpha}}(\beta_0)$.

For U_1, a first-order Taylor expansion around β_0 yields

$$\begin{aligned}
U_1 &= \frac{1}{\sqrt{N}} \Bigg\{ \sum_{i=1}^{n} \sum_{j>i} \frac{\partial \Delta_{ij}}{\partial \beta^\top}(\widehat{\alpha}_i(\bar{\beta}), \widehat{\alpha}_j(\bar{\beta}), \bar{\beta}) \\
&\qquad + \sum_{i=1}^{n} \sum_{j\neq i} \frac{\partial \Delta_{ij}}{\partial \alpha_i}(\widehat{\alpha}_i(\bar{\beta}), \widehat{\alpha}_j(\bar{\beta}), \bar{\beta}) \frac{\partial \widehat{\alpha}_i}{\partial \beta^\top}(\bar{\beta}) \Bigg\} (\widehat{\beta} - \beta_0) \\
&= \left\{ \Delta_\beta(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta})^\top - \Delta_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta})^\top \mathbf{J}_{11}(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta})^{-1} \mathbf{J}_{12}(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta}) \right\} \\
&\qquad \times \sqrt{N}(\widehat{\beta} - \beta_0) \\
&= (\Delta_{\beta,0}^\top - \Delta_{\boldsymbol{\alpha},0}^\top \mathbf{J}_{11,0}^{-1} \mathbf{J}_{12,0}) \sqrt{N}(\widehat{\beta} - \beta_0) + o_p(1),
\end{aligned} \tag{29}$$

where $\bar{\beta}$ lies in the segment between $\widehat{\beta}$ and β_0 and the last equality uses the fact that $\bar{\beta} \xrightarrow{p} \beta_0$ and $\left\|\widehat{\boldsymbol{\alpha}}(\bar{\beta}) - \boldsymbol{\alpha}_0\right\|_\infty \xrightarrow{p} 0$.

For U_2, a third-order Taylor expansion yields the following, with $\boldsymbol{\zeta} = \widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0$ as before:

$$\begin{aligned}
U_2 &= \sqrt{N} \Delta_{\boldsymbol{\alpha}}^\top \boldsymbol{\zeta} + \frac{1}{2\sqrt{N}} \sum_{k=1}^{n} \zeta_k \sum_{i=1}^{n} \sum_{j>i} \frac{\partial^2 \Delta_{ij}(\boldsymbol{\alpha}_0, \beta_0)}{\partial \alpha_k \partial \boldsymbol{\alpha}^\top} \boldsymbol{\zeta} \\
&\quad + \frac{1}{6\sqrt{N}} \sum_{k=1}^{n} \sum_{l=1}^{n} \zeta_k \zeta_l \sum_{i=1}^{n} \sum_{j>i} \frac{\partial^3 \Delta_{ij}(\bar{\boldsymbol{\alpha}}, \beta_0)}{\partial \alpha_k \partial \alpha_l \partial \boldsymbol{\alpha}^\top} \boldsymbol{\zeta}.
\end{aligned} \tag{30}$$

Recall that $\mathbf{R}_k^\mu$ and $\mathbf{G}_k^\mu$ are defined analogously to $\mathbf{R}_k$ and $\mathbf{G}_k$ in Appendix A. Similarly to the proof of Theorem 1, we can show that the second term of (30) converges in probability to a bias term $\lim_{n\to\infty}\frac{1}{2\sqrt{N}}\mathrm{Tr}\left[\mathbf{J}_{11,0}^{-1}\mathbf{V}_{11,0}\left(\mathbf{J}_{11,0}^{-1}\right)^\top\mathbf{R}_{k,0}^\mu\right]$ in (17). The last term of (30) is $o_p(1)$ (equivalent to the limit of part (IV) of (21)). By Lemma 8(b),

$$\sqrt{N}e_k^\top\Delta_{\boldsymbol{\alpha},0}^\top\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0 \xrightarrow{p} \lim_{n\to\infty}\frac{1}{2\sqrt{N}}\mathrm{Tr}[\mathbf{J}_{11,0}^{-1}\mathbf{V}_{11,0}\mathbf{J}_{11,0}^{-1}\mathbf{G}_{k,0}^\mu].$$

Additionally, from the proof of Theorem 1, we have

$$\sqrt{N}(\widehat{\beta}-\beta_0) = -\mathrm{J}_0^{-1}\{\frac{1}{\sqrt{N}}m_{2,0}-\frac{1}{\sqrt{N}}\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0}-B_0\}+o_p(1). \tag{31}$$

By Lemma 8(c),

$$\left\|\boldsymbol{\zeta}+\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0}+\frac{1}{2}\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0\right\|_\infty = O_p((\log n)^{3/2}/n^{3/2}). \tag{32}$$

Substituting (31) and (32) into (29) and (30) respectively, we have

$$\begin{aligned}
&\sqrt{N}(\widehat{\delta}-\bar{\Delta}_n)-B_\beta-B_\alpha \\
= &-(\Delta_{\beta,0}^\top-\Delta_{\boldsymbol{\alpha},0}^\top\mathbf{J}_{11,0}^{-1}\mathbf{J}_{12,0})\mathrm{J}_0^{-1}\left\{\frac{1}{\sqrt{N}}m_{2,0}-\frac{1}{\sqrt{N}}\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0}\right\} \\
&-\sqrt{N}\Delta_{\boldsymbol{\alpha},0}^\top\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0}+o_p(1) \\
= &-\frac{1}{\sqrt{N}}(\Delta_{\beta,0}^\top-\Delta_{\boldsymbol{\alpha},0}^\top\mathbf{J}_{11,0}^{-1}\mathbf{J}_{12,0})\mathrm{J}_0^{-1}m_{2,0} \\
&+\frac{1}{\sqrt{N}}\left[(\Delta_{\beta,0}^\top-\Delta_{\boldsymbol{\alpha},0}^\top\mathbf{J}_{11,0}^{-1}\mathbf{J}_{12,0})\mathrm{J}_0^{-1}\mathbf{J}_{21,0}-N\Delta_{\boldsymbol{\alpha},0}^\top\right]\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0}+o_p(1),
\end{aligned}$$

where B_β is defined in (17).

Finally, by the Lindeberg-Feller CLT, we have

$$\sqrt{N}(\widehat{\delta}-\bar{\Delta}_n)-B_\beta-B_\alpha \xrightarrow{d} \mathcal{N}(0,\Sigma_\Delta). \tag{33}$$

Combining (28), (33), and the fact that $\widehat{\delta}-\bar{\Delta}_n$ is uncorrelated with $\bar{\Delta}_n-\delta_0$ asymptotically, we have

$$\left(\frac{\Sigma_\Delta}{N}+\frac{4\Sigma_\delta}{n}\right)^{-1/2}\left(\widehat{\delta}-\delta_0-\frac{1}{\sqrt{N}}B_\beta-\frac{1}{\sqrt{N}}B_\alpha\right)\xrightarrow{d}\mathcal{N}(0,I_K).$$

Since the asymptotic normality of the plug-in estimator $\widehat{\delta}$ has already been established, the asymptotic normality of $\widehat{\delta}_{\mathrm{BG}}$ follows by an argument analogous to the proof of Theorem 3, and is therefore omitted for brevity. $\square$

References

Bonhomme, S. (2012). Functional differencing. *Econometrica 80*(4), 1337–1385.

Bonhomme, S., K. Jochmans, and M. Weidner (2024). A Neyman-orthogonalization approach to the incidental parameter problem. arXiv preprint arXiv:2412.10304.

Breiman, L. (1996). Bagging predictors. *Machine Learning 24*(2), 123–140.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2, pp. 1247–1318. North-Holland.

Chatterjee, S., P. Diaconis, and A. Sly (2011). Random graphs with given degree sequence. *Annals of Applied Probability 21*(4), 1400–1435.

Chen, M., I. Fernández-Val, and M. Weidner (2021). Nonlinear factor models for network and panel data. *Journal of Econometrics 220*(2), 296–324.

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014). Local identification of nonparametric and semiparametric models. *Econometrica 82*(2), 785–809.

de Paula, Á. (2020a). Econometric models of network formation. *Annual Review of Economics 12*, 775–799.

de Paula, Á. (2020b). Trade networks and the strength of strong ties. *Advances in Econometrics 42*.

de Paula, Á., S. Richards-Shubik, and E. Tamer (2018). Identifying preferences in networks with bounded degree. *Econometrica 86*(1), 263–288.

De Weerdt, J. (2004). Risk-sharing and endogenous network formation. In *Insurance Against Poverty*. Oxford University Press.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*(3), 991–1030.

Dzemski, A. (2019). An empirical model of dyadic link formation in network with unobserved heterogeneity. *Review of Economics and Statistics 101*(5), 763–776.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics 192*(1), 291–312.

Fernández-Val, I. and M. Weidner (2018). Fixed effects estimation of large-T panel data models. *Annual Review of Economics 10*(1), 109–138.

Gao, W. Y. (2020). Nonparametric identification in index models of link formation. *Journal of Econometrics 215*(2), 399–413.

Gao, W. Y., M. Li, and S. Xu (2023). Logical differencing in dyadic network formation models with nontransferable utilities. *Journal of Econometrics 235*(1), 302–324.

Gao, W. Y., M. Li, and Z. Xu (2026). Tractable identification of strategic network formation models with unobserved heterogeneity. Working paper.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Graham, B. S. (2020). Network data. In *Handbook of Econometrics*, Volume 7, pp. 111–218. Elsevier.

Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*(4), 1295–1319.

Hirano, K. and J. H. Wright (2017). Forecasting with model uncertainty: Representations and risk reduction. *Econometrica 85*(2), 617–643.

Honoré, B. E. and Á. de Paula (2021). Identification in simple binary outcome panel data models. *Econometrics Journal 24*(2).

Hughes, D. W. (2026). A jackknife bias correction for nonlinear network data models with fixed effects. *Journal of Econometrics 253*, 106130.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory 71*(1), 44–74.

Kinnan, C., K. Samphantharak, R. Townsend, and D. Vera-Cossio (2024). Propagation and insurance in village networks. *American Economic Review 114*(1), 252–284.

Le Cam, L. M. (1969). *Théorie Asymptotique de la Décision Statistique*. Presses de l'Université de Montréal.

Liao, C., Z. Mei, and Z. Shi (2024). Nickell meets Stambaugh: A tale of two biases in panel predictive regressions. *arXiv preprint arXiv:2410.09825*.

Mei, Z., L. Sheng, and Z. Shi (2026). Nickell bias in panel local projection: Financial crises are worse than you think. *Journal of International Economics*, 104210.

Qu, L., L. Chen, T. Yan, and Y. Chen (2025). Inference in semiparametric formation models for directed networks. *Journal of Business & Economic Statistics*. Forthcoming.

Rao, C. R. (1992). Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in Statistics: Foundations and Basic Theory*, pp. 235–247. Springer.

Shi, Z. and X. Chen (2016). A structural pairwise network model with individual heterogeneity. Available at SSRN 2848375.

van der Vaart, A. W. (2000). *Asymptotic Statistics*, Volume 3. Cambridge University Press.

Yan, T. (2019). Approximating the inverse of a diagonally dominant matrix with positive elements. arXiv Preprint arXiv:1902.00668.

Yan, T., B. Jiang, S. E. Fienberg, and C. Leng (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association 114*(526), 857–868.

Yan, T., H. Qin, and H. Wang (2016). Asymptotics in undirected random graph models parameterized by the strengths of vertices. *Statistica Sinica 26*(1), 273–293.

Zeleneev, A. (2026). Identification and estimation of network models with nonparametric unobserved heterogeneity. arXiv preprint arXiv:2602.06885.

Supplement to “Bagging the Network”*

Ming Li[a], Zhentao Shi[b], and Yapeng Zheng[b]

[a]Department of Economics and Risk Management Institute, National University of Singapore

[b]Department of Economics, The Chinese University of Hong Kong

This supplement contains material that complements the main text. Section A provides proofs of the supporting lemmas stated in the appendix and formalizes the multi-network extension of Remark 6 (Section A.7). Section B develops the full analysis of link function misspecification, including the identification assumption, formal theorems, and proofs. Section C reports extended Monte Carlo simulation results: the non-concavity challenge of direct MLE, extended NTU results on fixed-effect recovery, APE estimation, link-function misspecification, and sparser networks, and a brief summary of analogous TU results. Section D provides detailed data descriptions and summary statistics for both empirical applications.

A Proofs of Supporting Lemmas

We write “$a_n \asymp b_n$” to denote $a_n = O(b_n)$ and $b_n = O(a_n)$, and use $C_1, C_2, \ldots$ for positive finite constants. All probabilities are conditional on $\boldsymbol{\alpha}$ and $\mathbf{x}$ unless stated otherwise.

A.1 Inverse Approximation and Diagonal Bounds

We adapt Theorem 1 of Yan (2019) to the NTU framework to analytically approximate the inverse of the Jacobian matrix $\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)$ and bound the approximation errors. Similar techniques have been used to prove asymptotic normality in network estimation problems; see, for example, Yan and Xu (2013), Graham (2017), and Yan, Jiang, Fienberg, and Leng (2019). We prove that $\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)$ is non-singular for n large enough and $\mathbf{J}_{11}^{-1}(\boldsymbol{\alpha}, \beta)$ is well approximated by a diagonal matrix.

*Li: `mli@nus.edu.sg`. Shi (corresponding author): `zhentao.shi@cuhk.edu.hk`, 9F Esther Lee Building, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China. Zheng: `yapengzheng@link.cuhk.edu.hk`.

Proof of Lemma 2. Following Yan (2019), let I_n be the $n \times n$ identity matrix and define $\mathbf{F} = (f_{ij})_{n\times n} = \mathbf{A}^{-1} - \mathbf{B}$, $\mathbf{U} = I_n - \mathbf{AB}$, $\mathbf{W} = \mathbf{BU}$, which satisfy $\mathbf{F} = \mathbf{FU} + \mathbf{W}$. Direct calculation gives $u_{ij} = (\delta_{ij} - 1)a_{ij}/a_{jj}$ and $w_{ij} = (\delta_{ij} - 1)a_{ij}/(a_{ii}a_{jj})$, so that $\max(|w_{ij}|, |w_{ij} - w_{ik}|) \leq M/[m^2(n-1)^2]$ for all i, j, k.

The recursion $f_{ij} = \sum_k f_{ik}(\delta_{kj} - 1)a_{kj}/a_{jj} + w_{ij}$ and the identity $1 \equiv \sum_{k\neq\theta} a_{k\theta}/a_{\theta\theta} + \Delta_\theta/a_{\theta\theta}$ yield, after the algebraic manipulation detailed in equations (13)–(17) of Yan (2019) (obtained by maximizing over all cardinalities λ of the index set),

$$f_{i\theta} - f_{i\xi} \leq \frac{M/[m^2(n-1)^2]}{C(n,m,M)},$$

where $f_{i\theta} \coloneqq \max_k f_{ik}$, $f_{i\xi} \coloneqq \min_k f_{ik}$, and $C(n,m,M) \asymp 1$ when $m/M \asymp 1$. Moreover, for each fixed i,

$$\sum_{k=1}^{n} f_{ik}a_{ki} = \sum_{k=1}^{n} \left([\mathbf{A}^{-1}]_{ik} - \frac{\delta_{ik}}{a_{ii}} \right) a_{ki} = 1 - 1 = 0,$$

so $f_{i\xi} < 0 < f_{i\theta}$ and therefore $\max_k |f_{ik}| \leq f_{i\theta} - f_{i\xi}$. We thus obtain $\|\mathbf{F}\|_{\max} = O(n^{-2})$. $\square$

Based on Lemma 2, we prove that $\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)$ is non-singular for $(\boldsymbol{\alpha}, \beta) \in \mathbb{A} \times \mathbb{B}$ and large n.

Lemma A.1. *If Assumptions 1, 2, and 4 hold, for n large enough, the Jacobian matrix $\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)$ is invertible for all $(\boldsymbol{\alpha}, \beta) \in \mathbb{A} \times \mathbb{B}$.*

Proof. We partition $\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)$ into a block matrix as

$$\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta) = \begin{pmatrix} [\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{(1:n-1)\times(1:n-1)} & [\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{(1:n-1)\times n} \\ [\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{n\times(1:n-1)} & [\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{nn} \end{pmatrix},$$

where the subscript denotes the specific rows/columns that each sub-matrix includes. Recall that $[\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{ii} = \sum_{j\neq i} [\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{ji}$. The first sub-matrix $[\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{(1:n-1)\times(1:n-1)}$ is strictly diagonally dominant with all negative entries, hence it is non-singular. Lemma 2 demonstrates that its inverse can be approximated by $\operatorname{diag}\left([\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{11}^{-1}, \ldots, [\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{n-1n-1}^{-1}\right)$ with maximum entry-wise error of $O(n^{-2})$. Under Assumptions 2 and 4, $[\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{ii} \asymp -n$, $[\mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)]_{ij} \asymp -1$, $j \neq i$. Write $\mathbf{J} \coloneqq \mathbf{J}_{11}(\boldsymbol{\alpha}, \beta)$ and partition it as $\mathbf{J} = \left(\begin{smallmatrix} \mathbf{J}_{1:n-1,1:n-1} & \mathbf{J}_{1:n-1,n} \\ \mathbf{J}_{n,1:n-1} & \mathbf{J}_{nn} \end{smallmatrix}\right)$. Then

$$\mathbf{J}_{n,1:n-1}\mathbf{J}_{1:n-1,1:n-1}^{-1}\mathbf{J}_{1:n-1,n}$$

$$
\begin{aligned}
&= \mathbf{J}_{n,1:n-1}\mathrm{diag}\big(\mathbf{J}_{11}^{-1},\ldots,\mathbf{J}_{n-1\,n-1}^{-1}\big)\mathbf{J}_{1:n-1,n} \\
&\quad + O(n^{-2}) \times \mathbf{J}_{n,1:n-1}\mathbf{1}\mathbf{1}^\top\mathbf{J}_{1:n-1,n} \\
&= \sum_{i\neq n}\frac{\mathbf{J}_{ni}\mathbf{J}_{in}}{\sum_{j\neq i}\mathbf{J}_{ji}} + O(n^{-2})\Big(\sum_{i\neq n}\mathbf{J}_{ni}\Big)\Big(\sum_{i\neq n}\mathbf{J}_{in}\Big) = O(1).
\end{aligned}
$$

Thus, the Schur complement satisfies $\mathbf{J}_{nn}-\mathbf{J}_{n,1:n-1}\mathbf{J}_{1:n-1,1:n-1}^{-1}\mathbf{J}_{1:n-1,n} \asymp -n-O(1) \neq 0$ for n large enough. By the block-determinant formula,

$$
\det(\mathbf{J}) = \det(\mathbf{J}_{1:n-1,1:n-1}) \times \big(\mathbf{J}_{nn} - \mathbf{J}_{n,1:n-1}\mathbf{J}_{1:n-1,1:n-1}^{-1}\mathbf{J}_{1:n-1,n}\big) \neq 0
$$

for n large enough. Hence, $\mathbf{J}_{11}(\boldsymbol{\alpha},\beta)$ is invertible for large n. $\square$

For the inverse of $\mathbf{J}_{11}(\boldsymbol{\alpha},\beta)$, it is straightforward to verify that $-\mathbf{J}_{11}(\boldsymbol{\alpha},\beta)$ satisfies conditions in Lemma 2. Let $\mathbf{T}(\boldsymbol{\alpha},\beta) = [\mathrm{diag}\,(\mathbf{J}_{11}(\boldsymbol{\alpha},\beta))]^{-1}$. Applying Lemma 2 to $-\mathbf{J}_{11}(\boldsymbol{\alpha},\beta)$, we have $\|[-\mathbf{J}_{11}(\boldsymbol{\alpha},\beta)]^{-1} + \mathbf{T}(\boldsymbol{\alpha},\beta)\|_{\max}= O(n^{-2})$ under Assumptions 1–4. All of these results could be applied to $\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}},\boldsymbol{\alpha})$, where $\mathbf{T}^{\circ}(\widehat{\boldsymbol{\alpha}},\boldsymbol{\alpha})$ denotes the diagonal approximation for $[\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}},\boldsymbol{\alpha})]^{-1}$. The diagonal approximation technique extends from $\mathbf{J}_{11}$ to the information matrix $\mathbf{I}_{11}$, and combined with dyadic locality, yields the profiling-weight bounds used in the proof of Theorem 2.

Let $v_k(\boldsymbol{\alpha},\beta) \coloneqq \mathbf{I}_{12}(\boldsymbol{\alpha},\beta)e_k$ denote the kth column of $\mathbf{I}_{12}(\boldsymbol{\alpha},\beta)$, so that $w_k(\boldsymbol{\alpha},\beta) = \mathbf{I}_{11}(\boldsymbol{\alpha},\beta)^{-1}v_k(\boldsymbol{\alpha},\beta)$.

Proof of Lemma 3. By Assumption 4, the information kernels satisfy $[\mathbf{I}_{11}]_{ij} \asymp 1$ for $i \neq j$ (cf. Appendix A), so $\mathbf{D}_{ii} = [\mathbf{I}_{11}]_{ii} \asymp n$ uniformly. Write $\mathbf{I}_{11} = \mathbf{D}^{1/2}\mathbf{Q}\mathbf{D}^{1/2}$ where $\mathbf{Q} = \mathbf{D}^{-1/2}\mathbf{I}_{11}\mathbf{D}^{-1/2}$. Then $\mathbf{I}_{11}^{-1} = \mathbf{D}^{-1/2}\mathbf{Q}^{-1}\mathbf{D}^{-1/2}$, whence

$$
\|\mathbf{I}_{11}^{-1}\|_\infty \leq \|\mathbf{D}^{-1/2}\|_\infty^2\|\mathbf{Q}^{-1}\|_\infty = O(n^{-1})
$$

by Assumption 6. For the entrywise bound, note that $\mathbf{Q}_{ii} = 1$ and, for $i \neq j$, $|\mathbf{Q}_{ij}| = |[\mathbf{I}_{11}]_{ij}|/\sqrt{\mathbf{D}_{ii}\mathbf{D}_{jj}} = O(n^{-1})$. Using the identity $\mathbf{Q}^{-1} - I_n = -(\mathbf{Q} - I_n)\mathbf{Q}^{-1}$,

$$
\|\mathbf{Q}^{-1} - I_n\|_{\max} \leq \|\mathbf{Q} - I_n\|_{\max} \cdot \|\mathbf{Q}^{-1}\|_1 = O(n^{-1}).
$$

Consequently, $\mathbf{I}_{11}^{-1} - \mathbf{D}^{-1} = \mathbf{D}^{-1/2}(\mathbf{Q}^{-1} - I_n)\mathbf{D}^{-1/2}$, so

$$
\|\mathbf{I}_{11}^{-1} - \mathbf{D}^{-1}\|_{\max} \leq \|\mathbf{Q}^{-1} - I_n\|_{\max} \cdot \max_i \mathbf{D}_{ii}^{-1} = O(n^{-1}) \cdot O(n^{-1}) = O(n^{-2}). \ \square
$$

Lemma A.2 (Dyadic locality)**.** *If Assumptions 1, 2, and 4 hold, then uniformly over*

$(\boldsymbol{\alpha}, \beta) \in \mathbb{A} \times \mathbb{B}$ and all $k \in \{1, \ldots, K\}$,

$$\|v_k(\boldsymbol{\alpha}, \beta)\|_\infty = O(n) \text{ and } \left\| \frac{\partial v_k(\boldsymbol{\alpha}, \beta)}{\partial \beta} \right\|_\infty = O(n).$$

Moreover, for every $j \in \{1, \ldots, n\}$,

$$\left| \left[\frac{\partial v_k}{\partial \alpha_j} \right]_j \right| = O(n) \text{ and } \sup_{i \neq j} \left| \left[\frac{\partial v_k}{\partial \alpha_j} \right]_i \right| = O(1).$$

Further, for any deterministic vector $u \in \mathbb{R}^n$ with $\|u\|_\infty \leq C$,

$$\left\| \frac{\partial \mathbf{I}_{11}}{\partial \beta} u \right\|_\infty = O(n),$$

and, for each j,

$$\left| \left[\frac{\partial \mathbf{I}_{11}}{\partial \alpha_j} u \right]_j \right| = O(n) \text{ and } \sup_{i \neq j} \left| \left[\frac{\partial \mathbf{I}_{11}}{\partial \alpha_j} u \right]_i \right| = O(1).$$

Proof. Since $[v_k]_i = [\mathbf{I}_{12}]_{ik} = \sum_{l \neq i} \frac{p_{il}^{(1,0,0)} p_{il}^{(0,0,1)} x_{il,k}}{p_{il}(1-p_{il})}$, Assumption 4 gives $\|v_k\|_\infty = O(n)$. Differentiating with respect to β and using the bound on third-order derivatives of p yields $\|\partial_\beta v_k\|_\infty = O(n)$.

Fix j. If $i = j$, every term in $[v_k]_j = \sum_{l \neq j} (\cdot)$ depends on α_j, so $|[\partial_{\alpha_j} v_k]_j| \leq \sum_{l \neq j} |\partial_{\alpha_j} (\cdot)| = O(n)$. If $i \neq j$, only the single summand $l = j$ in $[v_k]_i$ depends on α_j, giving $|[\partial_{\alpha_j} v_k]_i| = O(1)$.

For $\partial_\beta \mathbf{I}_{11}$, the ith component of $(\partial_\beta \mathbf{I}_{11})u$ sums n bounded terms (by Assumption 4) times $\|u\|_\infty$, yielding $O(n)$. For $\partial_{\alpha_j} \mathbf{I}_{11}$, dyadic locality implies that when $i = j$ both $\partial_{\alpha_j} [\mathbf{I}_{11}]_{jj}$ and the off-diagonal derivatives $\partial_{\alpha_j} [\mathbf{I}_{11}]_{j\ell}$ for $\ell \neq j$ contribute, giving a total of $O(n)$. When $i \neq j$, at most two entries ($[\mathbf{I}_{11}]_{ii}$ through the $l = j$ summand, and $[\mathbf{I}_{11}]_{ij}$) have nonzero α_j-derivatives, each bounded by Assumption 4, so the contribution is $O(1)$. $\square$

Proof of Lemma 4. Since $w_k = \mathbf{I}_{11}^{-1} v_k$, Lemmas 3 and A.2 give $\|w_k\|_\infty \leq \|\mathbf{I}_{11}^{-1}\|_\infty \|v_k\|_\infty = O(n^{-1}) \cdot O(n) = O(1)$.

Differentiating $\mathbf{I}_{11} w_k = v_k$ with respect to a generic parameter $\theta \in \{\beta, \alpha_1, \ldots, \alpha_n\}$,

$$\partial_\theta w_k = \mathbf{I}_{11}^{-1} \underbrace{(\partial_\theta v_k - (\partial_\theta \mathbf{I}_{11}) w_k)}_{r_{k,\theta}}.$$

Case $\theta = \beta$. By Lemma A.2 and $\|w_k\|_\infty = O(1)$, $\|r_{k,\beta}\|_\infty = O(n)$. Hence $\|\partial_\beta w_k\|_\infty \leq \|\mathbf{I}_{11}^{-1}\|_\infty \|r_{k,\beta}\|_\infty = O(n^{-1}) \cdot O(n) = O(1)$.

Case $\theta = \alpha_j$. By Lemma A.2 and $\|w_k\|_\infty = O(1)$,

$$|[r_{k,j}]_j| = O(n), \qquad \sup_{i \neq j} |[r_{k,j}]_i| = O(1), \qquad \|r_{k,j}\|_1 = O(n).$$

Decomposing via the diagonal approximation in Lemma 3,

$$[\partial_{\alpha_j} w_k]_i = \mathbf{D}_{ii}^{-1}[r_{k,j}]_i + [(\mathbf{I}_{11}^{-1} - \mathbf{D}^{-1})r_{k,j}]_i.$$

For $i = j$: $\mathbf{D}_{jj}^{-1}|[r_{k,j}]_j| + \|\mathbf{I}_{11}^{-1} - \mathbf{D}^{-1}\|_{\max}\|r_{k,j}\|_1 = O(n^{-1}) \cdot O(n) + O(n^{-2}) \cdot O(n) = O(1)$. For $i \neq j$: $\mathbf{D}_{ii}^{-1}|[r_{k,j}]_i| + \|\mathbf{I}_{11}^{-1} - \mathbf{D}^{-1}\|_{\max}\|r_{k,j}\|_1 = O(n^{-1}) \cdot O(1) + O(n^{-2}) \cdot O(n) = O(n^{-1})$. $\square$

Remark A.1. Under additive models $p(\alpha_i, \alpha_j, x_{ij}^\top \beta) = F(\alpha_i + \alpha_j + x_{ij}^\top \beta)$ with $F' > 0$, we have $p^{(1,0,0)} = p^{(0,1,0)} = F'$, which gives $[\mathbf{I}_{11}]_{ij} = [F'(\cdot)]^2/[p_{ij}(1 - p_{ij})] > 0$ for $i \neq j$ and $[\mathbf{I}_{11}]_{ii} = \sum_{j \neq i}[\mathbf{I}_{11}]_{ji}$. Lemma 2 implies that $\|\mathbf{I}_{11}^{-1} - \mathbf{D}^{-1}\|_{\max} = O(n^{-2})$ and in consequence,

$$\|\mathbf{Q}^{-1}\|_1 \leq \|\mathbf{D}^{1/2}\mathbf{D}^{-1}\mathbf{D}^{1/2}\|_1 + \|\mathbf{D}^{1/2}(\mathbf{I}_{11}^{-1} - \mathbf{D}^{-1})\mathbf{D}^{1/2}\|_1 = O(1).$$

Hence, Assumption 6 is automatically satisfied for additive logit, probit, and other smooth link functions.

A.2 Well-definedness of limiting objects

Proof of Lemma 5. Throughout this proof, $p_{ij,0}^{(r_1,r_2,r_3)} = p^{(r_1,r_2,r_3)}(\alpha_{i0}, \alpha_{j0}, X_{ij}^\top \beta_0)$ per the convention, and $X_{ij} = h(X_i, X_j)$ per Assumption 3. By Assumptions 2–4, these quantities are uniformly bounded and $p_{ij,0}^{(1,0,0)} \geq c_2 > 0$.

Part (a), existence of J_0 *and its invertibility.* Using the decomposition from the proof of Theorem 1,

$$N^{-1}\mathrm{J}_{n,0} = N^{-1}\mathrm{J}_{22,0} - N^{-1}\mathbf{J}_{21,0}\mathbf{T}\mathbf{J}_{12,0} - N^{-1}\mathbf{J}_{21,0}(\mathbf{J}_{11,0}^{-1} - \mathbf{T})\mathbf{J}_{12,0},$$

where $\mathbf{T} \coloneqq [\mathrm{diag}(\mathbf{J}_{11,0})]^{-1}$ satisfies $\|\mathbf{J}_{11,0}^{-1} - \mathbf{T}\|_{\max} = O(n^{-2})$ by Lemma 2. We handle the three terms in turn.

Term 1: $(N^{-1}\mathrm{J}_{22,0})_{kl} = -N^{-1}\sum_{i<j} p_{ij,0}^{(0,0,1)} X_{ij,k}X_{ij,l}$ is a U-statistic of order 2 with bounded symmetric kernel. By the law of large numbers for U-statistics (e.g., van der Vaart, 2000, Theorem 12.3),

$$(N^{-1}\mathrm{J}_{22,0})_{kl} \xrightarrow{p} -\mathbb{E}\big[p^{(0,0,1)}(A_1, A_2, h(X_1, X_2)^\top \beta_0)\, h(X_1, X_2)_k\, h(X_1, X_2)_l\big],$$

with $(A_1, X_1), (A_2, X_2) \sim \nu$ independent, a finite matrix.

Term 2: As shown in the proof of Theorem 1, $-N^{-1}\mathbf{J}_{21,0}\mathbf{T}\mathbf{J}_{12,0}$ admits the representation

$$(-N^{-1}\mathbf{J}_{21,0}\mathbf{T}\mathbf{J}_{12,0})_{kl} = \frac{2}{n}\sum_{i=1}^{n} \frac{\bar{A}_{i,k,n}\,\bar{B}_{i,l,n}}{\bar{C}_{i,n}},$$

where $\bar{A}_{i,k,n} := (n-1)^{-1}\sum_{j\neq i} p^{(1,0,0)}_{ij,0} X_{ij,k}$, $\bar{B}_{i,l,n} := (n-1)^{-1}\sum_{j\neq i} p^{(0,0,1)}_{ij,0} X_{ij,l}$, and $\bar{C}_{i,n} := (n-1)^{-1}\sum_{j\neq i} p^{(1,0,0)}_{ij,0}$. Conditional on (α_{i0}, X_i), the quantities $\{(\alpha_{j0}, X_j)\}_{j\neq i}$ are i.i.d. from ν, so by the conditional LLN

$$\bar{A}_{i,k,n} \xrightarrow{p} a_k(\alpha_{i0}, X_i) := \mathbb{E}\big[p^{(1,0,0)}(\alpha_{i0}, A', h(X_i, X')^\top \beta_0) h(X_i, X')_k \mid \alpha_{i0}, X_i\big],$$

and similarly $\bar{B}_{i,l,n} \xrightarrow{p} b_l(\alpha_{i0}, X_i)$ and $\bar{C}_{i,n} \xrightarrow{p} c(\alpha_{i0}, X_i)$, with $c(\alpha_{i0}, X_i) \geq c_2 > 0$ uniformly by Assumption 4. Because all integrands are uniformly bounded and continuous on the compact $\mathbb{A}\times\mathbb{X}_0\times\mathbb{A}\times\mathbb{X}_0$, a uniform law of large numbers (Hoeffding's inequality applied to the bounded summands combined with a covering argument over the compact index set) gives the uniform convergence

$$\max_{1\leq i\leq n}\left|\frac{\bar{A}_{i,k,n}\bar{B}_{i,l,n}}{\bar{C}_{i,n}} - \frac{a_k(\alpha_{i0}, X_i) b_l(\alpha_{i0}, X_i)}{c(\alpha_{i0}, X_i)}\right| \xrightarrow{p} 0.$$

Combined with the LLN over i.i.d. $\{(\alpha_{i0}, X_i)\}$,

$$(-N^{-1}\mathbf{J}_{21,0}\mathbf{T}\mathbf{J}_{12,0})_{kl} \xrightarrow{p} 2\,\mathbb{E}\left[\frac{a_k(A,X) b_l(A,X)}{c(A,X)}\right],$$

a finite number since the integrand is bounded above (by $\sup|a_k|\sup|b_l|/c_2 < \infty$).

Term 3 (residual): Write $-\mathbf{J}_{11,0} = \mathbf{D}+\mathbf{O}$ with $\mathbf{D} = \mathrm{diag}(D_i)$, $D_i = \sum_{j\neq i} p^{(1,0,0)}_{ij,0} \asymp n$, and $\mathbf{O}_{ij} = p^{(0,1,0)}_{ij,0}$ for $i\neq j$, $\mathbf{O}_{ii} = 0$. The Neumann expansion yields $\mathbf{J}_{11,0}^{-1} - \mathbf{T} = \mathbf{D}^{-1}\mathbf{O}\mathbf{D}^{-1} + \mathbf{E}$, where $\|\mathbf{E}\|_\infty = O(n^{-2})$ by Lemma 6. The $\mathbf{E}$-contribution to $-N^{-1}\mathbf{J}_{21,0}(\mathbf{J}_{11,0}^{-1} - \mathbf{T})\mathbf{J}_{12,0}$ is $O(n^{-1}) = o(1)$ by the same scale accounting as in the proof of Theorem 1. For the leading term, using $(\mathbf{J}_{21,0})_{ki}/D_i = -\bar{A}_{i,k,n}/\bar{C}_{i,n}$ and $(\mathbf{J}_{12,0})_{jl}/D_j = -\bar{B}_{j,l,n}/\bar{C}_{j,n}$,

$$\begin{aligned}
-N^{-1}(\mathbf{J}_{21,0}\mathbf{D}^{-1}\mathbf{O}\mathbf{D}^{-1}\mathbf{J}_{12,0})_{kl} &= -\frac{1}{N}\sum_{i\neq j}\frac{\bar{A}_{i,k,n}\bar{B}_{j,l,n}}{\bar{C}_{i,n}\bar{C}_{j,n}}\,p^{(0,1,0)}_{ij,0} \\
&\xrightarrow{p} -\mathbb{E}\left[\frac{a_k(A_1,X_1) b_l(A_2,X_2)}{c(A_1,X_1)c(A_2,X_2)}\,p^{(0,1,0)}_{12}\right],
\end{aligned}$$

a finite number, by the uniform convergence of $\bar{A}, \bar{B}, \bar{C}$ and the LLN over i.i.d.

$\{(\alpha_{i0}, X_i)\}$. Combining Terms 1–3 establishes the existence of J_0 as a finite matrix. Note from (6) and the chain rule that $\nabla_{\beta^\top}\bar{S}_n(\beta_0) = N^{-1}\mathrm{J}_{n,0}$, which is of full rank by Assumption 5. Therefore $\mathrm{J}_0 = \mathrm{plim}_{n\to\infty} N^{-1}\mathrm{J}_{n,0}$ is invertible.

Part (a), existence of I_0. The proof is identical to that for J_0, with $\mathbf{I}_{11,0}, \mathbf{I}_{12,0}, \mathrm{I}_{22,0}$ replacing $-\mathbf{J}_{11,0}, -\mathbf{J}_{12,0}, -\mathrm{J}_{22,0}$ and Lemma 3 replacing Lemma 2. The entrywise scaling and the bounds in Assumption 4 carry over unchanged.

Part (b), existence of B_0 *and* b_0. The trace expression for B_{k0} in (15) is analyzed by substituting $\mathbf{J}_{11,0}^{-1} = \mathbf{T} + (\mathbf{J}_{11,0}^{-1} - \mathbf{T})$ twice and using the $O(n^{-2})$ entrywise residual bound; this reduces the trace to that of $\mathbf{T}\mathbf{V}_{11,0}\mathbf{T}^\top(\mathbf{G}_{k,0} + \mathbf{R}_{k,0})$ plus $o(1)$, whose diagonal entries are bounded continuous functions of the node-level averages $\bar{A}, \bar{B}, \bar{C}$. The scaled trace therefore equals $n^{-1}\sum_{i=1}^n \phi_k(\alpha_{i0}, X_i) + o(1)$ for a bounded continuous ϕ_k, converging to $\mathbb{E}[\phi_k(A, X)]$ by the LLN. The argument for b_{k0} is identical, using the bounds on $\mathbf{W}_{k,0}$ from Lemma 4.

Part (c), positive definiteness of Ω_0 *and* I_0. We prove $\mathrm{I}_0 \succ 0$; the argument for Ω_0 is analogous. Suppose $v^\top \mathrm{I}_0 v = 0$ for some nonzero $v \in \mathbb{R}^K$. By the Frisch–Waugh–Lovell residualization from the discussion following Assumption 6,

$$v^\top \mathrm{I}_0 v = \lim_{n\to\infty} \frac{1}{N} \sum_{i<j} p_{ij,0}(1 - p_{ij,0})(v^\top x_{ij} - w_{i,v} - w_{j,v})^2 = 0,$$

where $w_{i,v}$ are the profiling weights for $v^\top\beta$. Since $p_{ij,0}(1 - p_{ij,0}) \geq c_1(1 - c_1) > 0$ by Assumption 4, this forces $v^\top x_{ij} = w_{i,v} + w_{j,v}$ in $L^2(\mu_2)$, i.e., $v^\top x_{ij}$ is additively separable in the limit, contradicting Assumption 5 since a non-degenerate covariate distribution guarantees variation in $\bar{S}_n(\beta_0 + tv)$ orthogonal to additive node components. Hence $v = 0$. $\square$

A.3 Convergence Bounds

Proof of Lemma 6. This is equivalent to proving $\|\mathbf{B}^\top\mathbf{A}^\top\|_\infty \leq 1 - \frac{2(n-2)}{n-1}\delta^2$, which is a direct application of Lemma 2.1 of Chatterjee, Diaconis, and Sly (2011). $\square$

A.4 Deviation Bounds

Lemma 7 provides a bound on the deviation of the weighted sum of centered Bernoulli random variables, $\sum_{j\neq i}\lambda_{ij}(y_{ij} - p_{ij,0})$, and is used extensively in the proofs.

Let $\{\lambda_{ij}\}_{i,j=1}^{n}$ denote a sequence of bounded constants that satisfy $\max_{i,j}|\lambda_{ij}| < C_1$. In what follows, we apply the mean value theorem for vector-valued functions in its integral form, as in Chatterjee, Diaconis, and Sly (2011). For example,

$$\begin{aligned}\mathbf{m}_1(\widehat{\boldsymbol{\alpha}}, \beta) - \mathbf{m}_1(\boldsymbol{\alpha}, \beta) &= \left[\int_0^1 \mathbf{J}_{11}(\boldsymbol{\alpha} + t(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}), \beta)\, dt\right](\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \\ &=: \mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}; \beta)(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}).\end{aligned}$$

We write $\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}; \beta)$ as $\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha})$ whenever there is no confusion, and other integral form Jacobian matrices are defined similarly.

Proof of Lemma 7. First, notice that $|\lambda_{ij}(y_{ij} - p_{ij,0})| < 2C_1$ because $y_{ij} - p_{ij,0} \in (-1, 1)$; in addition, y_{ij}'s are independent Bernoulli random variables with expectations $p_{ij,0}$. By Hoeffding's inequality (see Theorem 2.8 of Boucheron, Lugosi, and Massart, 2013) for the sum of bounded and independent random variables, we have

$$\Pr\left(\frac{1}{n-1}\left|\sum_{j\neq i}\lambda_{ij}(y_{ij} - p_{ij,0})\right| > t\right) \leq 2\exp\left(-\frac{(n-1)t^2}{2C_1^2}\right).$$

Letting $t = C_1\sqrt{6(n-1)^{-1}\log n}$, we obtain

$$\Pr\left(\frac{1}{n-1}\left|\sum_{j\neq i}\lambda_{ij}(y_{ij} - p_{ij,0})\right| > C_1\sqrt{\frac{6\log n}{n-1}}\right) \leq 2n^{-\frac{3(n-1)}{n-1}} = 2n^{-3}.$$

By Boole's inequality,

$$\Pr\left(\max_{1\leq i\leq n}\frac{1}{n-1}\left|\sum_{j\neq i}\lambda_{ij}(y_{ij} - p_{ij,0})\right| > C_1\sqrt{\frac{6\log n}{n-1}}\right) \leq n\cdot 2n^{-3} = 2n^{-2}. \ \square$$

Using Lemma 7, we can bound the estimation error of $\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0$, which guarantees that our moment estimator is consistent for $\boldsymbol{\alpha}_0$ when β_0 is known. This result can be strengthened to prove the second part of Theorem 1, which we do in Appendix C.

Lemma A.3. *If Assumptions 1, 2, and 4 hold, then we have*

$$\Pr\left\{\left|\frac{1}{N}\sum_{i=1}^{n}\sum_{j>i}\lambda_{ij}(y_{ij} - p_{ij,0})\right| > C_1\sqrt{\frac{2\log N}{N}}\right\} \leq \frac{4}{n(n-1)}.$$

Proof. Similar to the proof of Lemma 7, by Hoeffding's inequality, we have

$$\Pr\left(\frac{1}{N}\left|\sum_{i=1}^{n}\sum_{j>i}\lambda_{ij}(y_{ij} - p_{ij,0})\right| > t\right) \leq 2\exp\left(-\frac{Nt^2}{2C_1^2}\right).$$

Letting $t = C_1\sqrt{\frac{2\log N}{N}}$, we obtain

$$\Pr\left(\frac{1}{N}\left|\sum_{i=1}^{n}\sum_{j>i}\lambda_{ij}(y_{ij}-p_{ij,0})\right| > C_1\sqrt{\frac{2\log N}{N}}\right) \le 2N^{-1} = \frac{4}{n(n-1)}. \square$$

Proof of Lemma 8. Part (a). The rest of the proof is conditional on the following event, which happens with probability at least $1-2n^{-2}$ by Lemma 7:

$$\mathcal{E}_n := \left\{\max_{1\le i\le n}\frac{1}{n-1}\left|\sum_{j\ne i}(y_{ij}-p_{ij,0})\right| \le \sqrt{\frac{6\log n}{n-1}} = O\left(\sqrt{\frac{\log n}{n}}\right)\right\}.$$

Since $\mathbf{m}_1(\widehat{\boldsymbol{\alpha}}(\beta_0),\beta_0) = 0$ by definition, a first-order Taylor expansion of $\mathbf{m}_1$ around $\boldsymbol{\alpha}_0$ gives $\widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0 = -\widetilde{\mathbf{J}}^{-1}\mathbf{m}_{1,0}$, where $\widetilde{\mathbf{J}} := \mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}}(\beta_0),\boldsymbol{\alpha}_0)$ and its diagonal approximation (per Lemma 2) is $\widetilde{\mathbf{T}} := \mathbf{T}^{\circ}(\widehat{\boldsymbol{\alpha}}(\beta_0),\boldsymbol{\alpha}_0)$. By Lemma 2 and the triangle inequality,

$$\begin{aligned}\|\widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0\|_\infty &= \|\widetilde{\mathbf{J}}^{-1}\mathbf{m}_{1,0}\|_\infty \\ &\le \|\widetilde{\mathbf{T}}\|_\infty\|\mathbf{m}_{1,0}\|_\infty + \|\widetilde{\mathbf{J}}^{-1}-\widetilde{\mathbf{T}}\|_\infty\|\mathbf{m}_{1,0}\|_\infty.\end{aligned}$$

For the first part, $\widetilde{\mathbf{T}}$ is diagonal with entries $O(n^{-1})$ uniformly; by Lemma 7,

$$\|\widetilde{\mathbf{T}}\|_\infty\|\mathbf{m}_{1,0}\|_\infty = O(n^{-1})\cdot\max_{1\le i\le n}\left|\sum\nolimits_{j\ne i}(y_{ij}-p_{ij,0})\right| = O\Big(\sqrt{\log n/n}\Big).$$

For the second part, by Lemma 2, $\|\widetilde{\mathbf{J}}^{-1}-\widetilde{\mathbf{T}}\|_\infty \le n\|\widetilde{\mathbf{J}}^{-1}-\widetilde{\mathbf{T}}\|_{\max} = O(n^{-1})$, so

$$\|\widetilde{\mathbf{J}}^{-1}-\widetilde{\mathbf{T}}\|_\infty\|\mathbf{m}_{1,0}\|_\infty = O\Big(\sqrt{\log n/n}\Big).$$

Combining these two results, we have $\|\widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0\|_\infty = O\left(\sqrt{\log n/n}\right)$.

Parts (b) and (c). By the third-order Taylor expansion, which is also used in the proof of Lemma 6 of Graham (2017), we have

$$\begin{aligned}&\mathbf{m}_1(\widehat{\boldsymbol{\alpha}}(\beta_0),\beta_0)-\mathbf{m}_{1,0} \\ &= \mathbf{J}_{11,0}[\widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0] + \frac{1}{2}\left[\sum_{k=1}^{n}(\widehat{\alpha}_k(\beta_0)-\alpha_{k0})\frac{\partial\mathbf{J}_{11}(\boldsymbol{\alpha}_0,\beta_0)}{\partial\alpha_k}\right][\widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0] \\ &\quad + \frac{1}{6}\left[\sum_{k=1}^{n}\sum_{l=1}^{n}(\widehat{\alpha}_k(\beta_0)-\alpha_{k0})(\widehat{\alpha}_l(\beta_0)-\alpha_{l0})\frac{\partial^2\mathbf{J}_{11}(\bar{\boldsymbol{\alpha}}^{kl},\beta_0)}{\partial\alpha_k\partial\alpha_l}\right][\widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0], \qquad \text{(A.1)}\end{aligned}$$

where for each (k,l), $\bar{\boldsymbol{\alpha}}^{kl} = \boldsymbol{\alpha}_0 + s_{kl}(\widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0)$ for some $s_{kl}\in(0,1)$.

Recall the definition of $\boldsymbol{\eta}_0$ and define the $n\times 1$ vector

$$\boldsymbol{\rho}_0 := \left[\sum_{k=1}^{n}\sum_{l=1}^{n}(\widehat{\alpha}_k(\beta_0)-\alpha_{k0})(\widehat{\alpha}_l(\beta_0)-\alpha_{l0})\frac{\partial^2 \mathbf{J}_{11}(\bar{\boldsymbol{\alpha}}^{kl},\beta_0)}{\partial\alpha_k\partial\alpha_l}\right][\widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0].$$

Let $\delta_i := \widehat{\alpha}_i(\beta_0)-\alpha_{i0}$, by part (a), we have $\sup_i|\delta_i| = O_p(\sqrt{\log n/n})$. Because only the kth row and the kth column of $\mathbf{J}_{11}(\boldsymbol{\alpha},\beta)$ contain functions of α_k, by a direct calculation we write the entries of $\Lambda_k := \frac{\partial \mathbf{J}_{11}(\boldsymbol{\alpha}_0,\beta_0)}{\partial\alpha_k}$ as

$$(\Lambda_k)_{kl} = -p_{kl,0}^{(1,1,0)}\ (l\neq k),\quad (\Lambda_k)_{lk} = -p_{kl,0}^{(2,0,0)}\ (l\neq k),\quad (\Lambda_k)_{kk} = -\textstyle\sum_{p\neq k} p_{kp,0}^{(2,0,0)},$$

with all other entries equal to zero. Hence, let $\Lambda = \sum_{k=1}^{n}(\widehat{\alpha}_k(\beta_0)-\alpha_{k0})\frac{\partial \mathbf{J}_{11}(\boldsymbol{\alpha}_0,\beta_0)}{\partial\alpha_k}$ whose entries are

$$\begin{aligned}
\Lambda_{ij} &= -(\widehat{\alpha}_i(\beta_0)-\alpha_{i0})p_{ij,0}^{(1,1,0)} - (\widehat{\alpha}_j(\beta_0)-\alpha_{j0})p_{ji,0}^{(2,0,0)},\ i\neq j,\\
\Lambda_{ii} &= -(\widehat{\alpha}_i(\beta_0)-\alpha_{i0})\sum_{j\neq i} p_{ij,0}^{(2,0,0)} - \sum_{k\neq i}(\widehat{\alpha}_k(\beta_0)-\alpha_{k0})p_{ik,0}^{(1,1,0)}.
\end{aligned}$$

Then, the ith element of $\boldsymbol{\eta}_0$ can be calculated as

$$\begin{aligned}
\eta_{i,0} &= \Lambda_{ii}\cdot(\widehat{\alpha}_i(\beta_0)-\alpha_{i0}) + \sum_{j\neq i}\Lambda_{ij}\cdot(\widehat{\alpha}_j(\beta_0)-\alpha_{j0})\\
&= -\sum_{j\neq i} p_{ij,0}^{(2,0,0)}(\widehat{\alpha}_i(\beta_0)-\alpha_{i0})^2\\
&\quad - 2\sum_{j\neq i} p_{ij,0}^{(1,1,0)}(\widehat{\alpha}_i(\beta_0)-\alpha_{i0})(\widehat{\alpha}_j(\beta_0)-\alpha_{j0})\\
&\quad - \sum_{j\neq i} p_{ji,0}^{(2,0,0)}(\widehat{\alpha}_j(\beta_0)-\alpha_{j0})^2.
\end{aligned}$$

It is clear that $\|\boldsymbol{\eta}_0\|_\infty = O_p(\log n/n)$ by $\|\widehat{\boldsymbol{\alpha}}(\beta_0)-\boldsymbol{\alpha}_0\|_\infty = O_p(\sqrt{\log n/n})$. Recall the symmetric matrix $\mathbf{M}_{i,n}$ defined in (13). Then, we have $\eta_{i,0} = \boldsymbol{\delta}^\top\mathbf{M}_{i,n,0}\boldsymbol{\delta}$. Recall from Appendix A that $\varphi_{k,0}^\top = e_k^\top\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}$ and $\mathbf{G}_{k,0} = \sum_{i=1}^{n}\varphi_{k,i,0}\mathbf{M}_{i,n,0}$. Hence, $e_k^\top\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0 = \varphi_{k,0}^\top\boldsymbol{\eta}_0 = \sum_{i=1}^{n}\varphi_{k,i,0}\,\eta_{i,0} = \boldsymbol{\delta}^\top\mathbf{G}_{k,0}\boldsymbol{\delta}$, where $e_k\in\mathbb{R}^K$ is the kth canonical basis vector. Substituting $\boldsymbol{\delta} = -\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0}+\mathbf{r}_n$ with $\|\mathbf{r}_n\|_\infty = O_p(\log n/n)$, we have

$$\begin{aligned}
\frac{1}{\sqrt{N}}e_k^\top\mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0 &= \frac{1}{\sqrt{N}}\mathbf{m}_{1,0}^\top(\mathbf{J}_{11,0}^{-1})^\top\mathbf{G}_{k,0}\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0}\\
&\quad - \frac{2}{\sqrt{N}}\mathbf{m}_{1,0}^\top(\mathbf{J}_{11,0}^{-1})^\top\mathbf{G}_{k,0}\mathbf{r}_n + \frac{1}{\sqrt{N}}\mathbf{r}_n^\top\mathbf{G}_{k,0}\mathbf{r}_n.
\end{aligned}$$

By the definition of $\mathbf{G}_k$ in Appendix A and bounded second-order derivatives, $\mathbf{G}_{k,0}$

has uniformly bounded row sums, so the last two terms are $o_p(1)$. Therefore,

$$\begin{aligned}\frac{1}{2\sqrt{N}}e_k^\top \mathbf{J}_{21,0}\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0 &= \frac{1}{\sqrt{N}}\mathbf{m}_{1,0}^\top(\mathbf{J}_{11,0}^{-1})^\top \mathbf{G}_{k,0}\mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0} + o_p(1) \\ &\xrightarrow{p} \lim_{n\to\infty}\frac{1}{2\sqrt{N}}\mathrm{Tr}[\mathbf{J}_{11,0}^{-1}\mathbf{V}_{11,0}(\mathbf{J}_{11,0}^{-1})^\top \mathbf{G}_{k,0}] = B_{k0}^{(1)}.\end{aligned}$$

Next, uniformly for $i \in \mathcal{I}_n$, we have:

$$\begin{aligned}|\rho_{i,0}| &\le C\sum_{j\neq i}\big(|\delta_i|^3 + |\delta_i|^2|\delta_j| + |\delta_i||\delta_j|^2 + |\delta_j|^3\big) \\ &\le 8C(n-1)(\sup_i|\delta_i|)^3 = O\big((\log n)^{3/2}/\sqrt{n}\big)\,,\end{aligned}$$

by bounded third-order derivatives. Hence,

$$\|\mathbf{J}_{11,0}^{-1}\boldsymbol{\rho}_0\|_\infty \le \|\mathbf{J}_{11,0}^{-1}\|_\infty\|\boldsymbol{\rho}_0\|_\infty = O\left(\frac{1}{n}\right)\cdot O_p\left(\frac{(\log n)^{3/2}}{\sqrt{n}}\right) = O_p\left(\frac{(\log n)^{3/2}}{n^{3/2}}\right).$$

Finally, by (A.1), we have

$$\left\|\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0 + \mathbf{J}_{11,0}^{-1}\mathbf{m}_{1,0} + \frac{1}{2}\mathbf{J}_{11,0}^{-1}\boldsymbol{\eta}_0\right\|_\infty = \left\|\frac{1}{6}\mathbf{J}_{11,0}^{-1}\boldsymbol{\rho}_0\right\|_\infty = O_p\left(\frac{(\log n)^{3/2}}{n^{3/2}}\right). \ \square$$

A.5 Uniform Convergence of the Concentrated Moment Equation

Recall the concentrated moment equation and its population counterpart:

$$S_n(\beta) := N^{-1}m_2(\widehat{\boldsymbol{\alpha}}(\beta),\beta) \text{ and } \bar{S}_n(\beta) := N^{-1}\mathbb{E}[m_2(\boldsymbol{\alpha}(\beta),\beta)|\mathbf{x},\boldsymbol{\alpha}_0],$$

where $\boldsymbol{\alpha}(\beta)$ is the unique solution to $\mathbb{E}[\mathbf{m}_1(\boldsymbol{\alpha},\beta)|\mathbf{x},\boldsymbol{\alpha}_0] = 0$.

Proof of Lemma 9. By the mean-value theorem, we have $\widehat{\boldsymbol{\alpha}}(\beta) - \boldsymbol{\alpha}(\beta) = -[\mathbf{J}_{11}^\circ(\widehat{\boldsymbol{\alpha}}(\beta),\boldsymbol{\alpha}(\beta))]^{-1}\mathbf{m}_{1,0}$. Decomposing $S_n(\beta) - \bar{S}_n(\beta)$ via the diagonal approximation $\mathbf{T}^\circ = [\mathrm{diag}(\mathbf{J}_{11}^\circ)]^{-1}$ from Lemma 2 gives

$$\begin{aligned}S_n(\beta) - \bar{S}_n(\beta) = &\underbrace{N^{-1}\textstyle\sum_i\sum_{j>i}(y_{ij}-p_{ij,0})x_{ij}}_{R_1} - \underbrace{N^{-1}\mathbf{J}_{21}^\circ\mathbf{T}^\circ\mathbf{m}_{1,0}}_{R_2} \\ &+ \underbrace{N^{-1}\mathbf{J}_{21}^\circ[\mathbf{T}^\circ - (\mathbf{J}_{11}^\circ)^{-1}]\mathbf{m}_{1,0}}_{R_3}.\end{aligned}$$

By Lemmas A.3 and 7, $\|R_1\|_\infty = O_p(\sqrt{(\log N)/N})$. Since $\mathbf{T}^\circ$ is diagonal with entries $O(n^{-1})$ and $\|\mathbf{J}_{21}^\circ\|_{\max} = O(n)$, $\|R_2\|_\infty = O_p(\sqrt{(\log n)/n})$. For R_3, using $\|\mathbf{T}^\circ -$

$(\mathbf{J}_{11}^{\circ})^{-1}\|_{\max} = O(n^{-2})$, $\|R_3\|_\infty = O_p(\sqrt{(\log n)/n})$. All bounds hold uniformly in $\beta \in \mathbb{B}$. $\square$

A.6 One-Step Score Bounds

Let $s_n(\boldsymbol{\alpha}, \beta) = s_2(\boldsymbol{\alpha}, \beta) - \sum_{i=1}^n s_{1i}(\boldsymbol{\alpha}, \beta) w_i(\boldsymbol{\alpha}, \beta)$, where $w_i(\boldsymbol{\alpha}, \beta)$ is the ith column of $\mathbf{I}_{12}(\boldsymbol{\alpha}, \beta)^\top \mathbf{I}_{11}(\boldsymbol{\alpha}, \beta)^{-1}$. Taking derivatives, we have

$$\nabla_{\boldsymbol{\alpha}^\top} s_n = \mathbf{H}_{12}^\top - \mathbf{I}_{12}^\top \mathbf{I}_{11}^{-1} \mathbf{H}_{11} - \textstyle\sum_{i=1}^n s_{1i}\, \partial_{\boldsymbol{\alpha}^\top} w_i \text{ and}$$
$$\nabla_{\beta^\top} s_n = \mathrm{H}_{22} - \mathbf{I}_{12}^\top \mathbf{I}_{11}^{-1} \mathbf{H}_{12} - \textstyle\sum_{i=1}^n s_{1i}\, \partial_{\beta^\top} w_i,$$

where all matrices are evaluated at $(\boldsymbol{\alpha}, \beta)$.

Proof of Lemma 10. Each row of $\mathbf{H}_{12,0} + \mathbf{I}_{12,0}$ is a weighted sum of $(y_{ij} - p_{ij,0})$ terms, so by Lemma 8 and the continuous mapping theorem (CMT),

$$N^{-1/2}\, \|\mathbf{H}_{12}(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0) + \mathbf{I}_{12}(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0)\|_{\max} = o_p(1).$$

Similarly,

$$N^{-1/2} \left\| \mathbf{I}_{12}(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0)^\top \mathbf{I}_{11}(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0)^{-1} \left[\mathbf{H}_{11}(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0) + \mathbf{I}_{11}(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0)\right] \right\|_{\max} = o_p(1).$$

Combining these two bounds,

$$N^{-1/2} \|\mathbf{H}_{12}(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0)^\top - \mathbf{I}_{12}^\top \mathbf{I}_{11}^{-1} \mathbf{H}_{11}\|_{\max} = o_p(1), \tag{A.2}$$

evaluated at $(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0)$. From the exact mean-value representation in the proof of Lemma 8, $\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0 = -\left[\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}}(\beta_0), \boldsymbol{\alpha}_0)\right]^{-1} \mathbf{m}_{1,0}$. Lemma 2 implies that the diagonal entries of $[\mathbf{J}_{11}^{\circ}(\widehat{\boldsymbol{\alpha}}(\beta_0), \boldsymbol{\alpha}_0)]^{-1}$ are $O_p(n^{-1})$ and the off-diagonal entries are $O(n^{-2})$. Hence, for any deterministic unit vector $\mathbf{c}$, $\mathbf{c}^\top(\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0)$ is a weighted sum of conditionally independent centered dyad shocks with total squared weight $O(n^{-1})$, so $\sqrt{n}\mathbf{c}^\top(\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0) = O_p(1)$. Combining this with (A.2) yields

$$\frac{1}{\sqrt{N}} [\mathbf{H}_{12}^\top - \mathbf{I}_{12}^\top \mathbf{I}_{11}^{-1} \mathbf{H}_{11}](\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0) = o_p(1). \tag{A.3}$$

Next, similarly to the process of finding the bias term in the proof of Theorem 1, we have

$$-\frac{1}{\sqrt{N}} \sum_{i=1}^n s_{1i}(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0) \frac{\partial w_{ki}(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0)}{\partial \boldsymbol{\alpha}^\top} (\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0)$$

$$= \frac{1}{\sqrt{N}} \mathbf{s}_{1,0}^\top \mathbf{W}_{k,0} \mathbf{J}_{11,0}^{-1} \mathbf{m}_{1,0} + o_p(1),$$

where $[\mathbf{W}_k(\boldsymbol{\alpha}, \beta)]_{ij} = \frac{\partial w_{ki}(\boldsymbol{\alpha},\beta)}{\partial \alpha_j}$ and the equality holds by $\|\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0\|_\infty = o_p(1)$ and the CMT.

The asymptotic bias $b_0 := (b_{10}, \dots, b_{K0})^\top$ for the one-step estimator is defined as

$$b_{k0} = \lim_{n\to\infty} \frac{1}{\sqrt{N}} \mathrm{Tr}[\mathbf{J}_{11,0}^{-1} \mathrm{Cov}(\mathbf{m}_{1,0}, \mathbf{s}_{1,0}) \mathbf{W}_{k,0}], \quad k = 1, \dots, K, \tag{A.4}$$

where the entries of the $n \times n$ covariance matrix $\mathrm{Cov}(\mathbf{m}_{1,0}, \mathbf{s}_{1,0})$ are

$$\begin{aligned} [\mathrm{Cov}(\mathbf{m}_{1,0}, \mathbf{s}_{1,0})]_{ij} &= \frac{p_{ji,0}^{(1,0,0)} \mathrm{Var}(y_{ij})}{p_{ji,0}(1 - p_{ij,0})} = p_{ji,0}^{(1,0,0)}, \quad 1 \le i \ne j \le n, \\ [\mathrm{Cov}(\mathbf{m}_{1,0}, \mathbf{s}_{1,0})]_{ii} &= \sum_{k\ne i} \frac{p_{ik,0}^{(1,0,0)} \mathrm{Var}(y_{ik})}{p_{ik,0}(1 - p_{ik,0})} = \sum_{k\ne i} p_{ik,0}^{(1,0,0)}, \quad 1 \le i \le n. \end{aligned} \tag{A.5}$$

Next, we have

$$\begin{aligned} &\frac{1}{\sqrt{N}} \mathbf{s}_{1,0}^\top \mathbf{W}_{k,0} \mathbf{J}_{11,0}^{-1} \mathbf{m}_{1,0} = \frac{1}{\sqrt{N}} \mathrm{Tr}(\mathbf{J}_{11,0}^{-1} \mathbf{m}_{1,0} \mathbf{s}_{1,0}^\top \mathbf{W}_{k,0}) \\ &= \frac{1}{\sqrt{N}} \mathrm{Tr}[\mathbf{J}_{11,0}^{-1} \mathrm{Cov}(\mathbf{m}_{1,0}, \mathbf{s}_{1,0}) \mathbf{W}_{k,0}] \\ &\quad + \frac{1}{\sqrt{N}} \left\{ \mathrm{Tr}(\mathbf{J}_{11,0}^{-1} \mathbf{m}_{1,0} \mathbf{s}_{1,0}^\top \mathbf{W}_{k,0}) - \mathrm{Tr}[\mathbf{J}_{11,0}^{-1} \mathrm{Cov}(\mathbf{m}_{1,0}, \mathbf{s}_{1,0}) \mathbf{W}_{k,0}] \right\} \\ &= R_1 + R_2. \end{aligned} \tag{A.6}$$

Notice that $R_1 \to b_{k0}$ by definition. By the law of large numbers for U-statistics, $R_2 \xrightarrow{p} 0$ under the bounds in Lemma 4. By (A.3) and (A.6), we have $N^{-1/2} \nabla_{\boldsymbol{\alpha}^\top} s_n(\widehat{\boldsymbol{\alpha}}(\beta_0), \beta_0)(\widehat{\boldsymbol{\alpha}}(\beta_0) - \boldsymbol{\alpha}_0) \xrightarrow{p} b_0$, which proves Lemma 10(a).

For the second result of Lemma 10, the law of large numbers gives $N^{-1}[\mathrm{H}_{22} - \mathbf{I}_{12}^\top \mathbf{I}_{11}^{-1} \mathbf{H}_{12}] \xrightarrow{p} -\mathrm{I}_0$ at $(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta})$. Since $\|\partial_\beta w_i\|_\infty = O(1)$ by Lemma 4 and $N^{-1} \sum_i |s_{1i}| = o_p(1)$ by the law of large numbers,

$$N^{-1} \nabla_{\beta^\top} s_n(\widehat{\boldsymbol{\alpha}}(\bar{\beta}), \bar{\beta}) + \mathrm{I}_0 = o_p(1). \tag{A.7}$$

By (A.2) and an argument analogous to the proof of the first result, $N^{-1} \nabla_{\boldsymbol{\alpha}^\top} s_n \cdot \partial \widehat{\boldsymbol{\alpha}} / \partial \beta^\top = o_p(1)$, which combined with (A.7) proves Lemma 10(b). $\square$

A.7 Extension to Multiple Networks

We formalize the multi-network extension of Remark 6. Consider V independent networks (V fixed, $\min_v n_v \to \infty$) sharing a common slope β_0 but with network-specific node fixed effects $\boldsymbol{\alpha}_{0,v}$. The link function $p(\cdot)$ may follow either the TU or NTU specification.

Modified assumptions. Assumptions 1–4 and 6 hold within each network with constants uniform in v. Assumption 5 is imposed on the pooled concentrated moment

$$\bar{S}_n(\beta) \coloneqq N^{-1} \sum_{v=1}^{V} \sum_{i<j} \mathbb{E}\left[(y_{ij,v} - p_{ij,v}(\boldsymbol{\alpha}_v(\beta), \beta))\, x_{ij,v} \,\middle|\, \mathbf{x}, \boldsymbol{\alpha}_{0,v} \right],$$

where $\boldsymbol{\alpha}_v(\beta)$ profiles the network-v degree equations and $N \coloneqq \sum_v \binom{n_v}{2}$.

Proposition A.1 (Multi-network extension)**.** *Under the modified assumptions above, the conclusions of Theorems 1–3 extend with $N = \sum_v \binom{n_v}{2}$, pooled bias $B_0 = \sum_v B_{0,v}$, and pooled information $\mathrm{I}_0 = \sum_v \mathrm{I}_{0,v}$. In particular, $\sqrt{N}(\widehat{\beta}_{\mathrm{BG}} - \beta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{I}_0^{-1})$.*

Proof of Proposition A.1. Independence across networks makes the score, Jacobian, and information block-diagonal in the network index, so $\mathbf{I}_{11}$, $\mathbf{J}_{11}$ and their inverses inherit the diagonal-dominance bounds of Lemma 3 per network with constants uniform in v. The contraction map for $\boldsymbol{\alpha}_v(\beta)$ in Lemma 1 is network-internal, and the per-network bagging step yields the $\sqrt{N_v/4}$-rate result of Theorem 3. The cross-network average then inherits $\sqrt{N}$-normality by independence. The proofs of Theorems 1–3 carry over after replacing $\sum_{i<j}$ with $\sum_v \sum_{i<j}$ and N with $\sum_v \binom{n_v}{2}$. $\square$

B Link Function Misspecification

We study dyadic network formation under possible link function misspecification in this section. The analysis applies to TU and NTU models. Suppose researchers misspecify the link function to be $q(\cdot)$ which differs from $p(\cdot)$ at points with strictly positive probability measure. For a fixed n, we impose the following identification assumption to facilitate the analysis. Let $q_{ij}(\boldsymbol{\alpha}, \beta) \coloneqq q(\alpha_i, \alpha_j, x_{ij}^\top \beta)$ be the misspecified probability of linking between i and j.

Assumption 7 (Identification under Link Function Misspecification)**.** *For sufficiently large n, the normalized nonlinear function*

$$\widetilde{S}_n(\beta) \coloneqq N^{-1} \sum_{i=1}^{n} \sum_{j>i} \left[p_{ij,0} - q_{ij}(\boldsymbol{\alpha}(\beta), \beta)\right] x_{ij}$$

has a unique root β_{n*}, and satisfies

$$\inf_{\beta \in \mathbb{B}: \|\beta - \beta_{n*}\|_2 \geq \delta} \left\| \widetilde{S}_n(\beta) \right\|_2 > 0$$

for all $\delta > 0$, where $\boldsymbol{\alpha}(\beta)$ is the unique solution to the following system of equations

$$\left(\sum_{j \neq 1} p_{1j,0} - \sum_{j \neq 1} q_{1j}(\boldsymbol{\alpha}, \beta), \cdots, \sum_{j \neq n} p_{nj,0} - \sum_{j \neq n} q_{nj}(\boldsymbol{\alpha}, \beta) \right)^{\top} = 0.$$

Assumption 7 is the counterpart of Assumption 5 under a misspecified link function. Similarly to Lemma 1, the degree-matching system in Assumption 7 has a unique solution with high probability under mild conditions on $(\boldsymbol{\alpha}_0, \beta_0)$ and β. Thus, Assumption 7 identifies the homophily parameter under link function misspecification. Notice that β_{n*} depends on the true link function $p(\cdot)$, misspecified link function $q(\cdot)$, and the true parameter values. As a result, β_{n*} may vary with n. The following theorem shows that the JMM estimator based on the misspecified link function $q(\cdot)$ is centered at β_{n*} up to a bias, which the split-network jackknife procedure removes asymptotically. Let $\boldsymbol{\alpha}_* \coloneqq \boldsymbol{\alpha}(\beta_{n*})$ with $\boldsymbol{\alpha}(\cdot)$ satisfying Assumption 7.

In the next Theorem, J_*, B_*, and Ω_* are defined analogously to J_0, B_0, and Ω_0 in Theorem 1, with the pseudo values $(\boldsymbol{\alpha}_*, \beta_{n*})$ and the misspecified link function $q(\cdot)$. For the sandwich-form variance Ω_*, the sandwich "meat" remains governed by the true DGP through $\mathrm{Var}(y_{ij} \mid \mathbf{x}, \boldsymbol{\alpha}_0) = p_{ij,0}(1 - p_{ij,0})$. Furthermore, Ω_* can be consistently estimated by the plug-in sandwich formula.

Theorem B.1 (JMM Estimation under Link Function Misspecification)**.** *If Assumptions 1–4 and 7 hold, then*

$$\sqrt{N}(\widehat{\beta} - \beta_{n*}) - \mathrm{J}_*^{-1} B_* \xrightarrow{d} \mathcal{N}(0, \Omega_*).$$

Theorem B.1 shows that the JMM estimator remains consistent for the pseudo-true value β_{n*} (the unique solution to the pseudo-moment equations in Assumption 7) even when the link function is misspecified.

Under the link function misspecification, the one-step estimator becomes

$$\widehat{\beta}_{\mathrm{OS}} \coloneqq \widehat{\beta} - \mathrm{H}(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})^{-1} s_n(\widehat{\boldsymbol{\alpha}}, \widehat{\beta}), \tag{B.1}$$

with the JMM estimator $(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})$ substituted in. Note that

$$\mathrm{H}(\boldsymbol{\alpha}, \beta) \coloneqq \mathrm{H}_{22}(\boldsymbol{\alpha}, \beta) - \mathbf{H}_{12}(\boldsymbol{\alpha}, \beta)^\top \mathbf{H}_{11}(\boldsymbol{\alpha}, \beta)^{-1} \mathbf{H}_{12}(\boldsymbol{\alpha}, \beta), \tag{B.2}$$

is the concentrated Hessian matrix, and

$$s_n(\boldsymbol{\alpha}, \beta) \coloneqq s_2(\boldsymbol{\alpha}, \beta) - \mathbf{H}_{12}(\boldsymbol{\alpha}, \beta)^\top \mathbf{H}_{11}(\boldsymbol{\alpha}, \beta)^{-1} \mathbf{s}_1(\boldsymbol{\alpha}, \beta),$$

is the concentrated score function. To define the pseudo-true target of $\widehat{\beta}_{\mathrm{OS}}$, write $\bar{\mathbf{H}}_{ab}(\boldsymbol{\alpha}, \beta) \coloneqq \mathbb{E}[\mathbf{H}_{ab}(\boldsymbol{\alpha}, \beta) \mid \mathbf{x}, \boldsymbol{\alpha}_0]$ for the expected Hessian blocks, and let

$$\bar{\mathrm{H}}(\boldsymbol{\alpha}, \beta) \coloneqq \bar{\mathrm{H}}_{22} - \bar{\mathbf{H}}_{12}^\top \bar{\mathbf{H}}_{11}^{-1} \bar{\mathbf{H}}_{12}, \qquad \bar{s}_n(\boldsymbol{\alpha}, \beta) \coloneqq \mathbb{E}[s_2] - \bar{\mathbf{H}}_{12}^\top \bar{\mathbf{H}}_{11}^{-1} \mathbb{E}[\mathbf{s}_1],$$

denote the corresponding population concentrated Hessian and score, with arguments $(\boldsymbol{\alpha}, \beta)$ suppressed for brevity. Under the link function misspecification, $\widehat{\beta}_{\mathrm{OS}}$ in (B.1) centers on the deterministic sequence

$$\beta_{n\sharp} \coloneqq \beta_{n*} - N^{-1} \mathrm{H}_*^{-1} \bar{s}_n(\boldsymbol{\alpha}_*, \beta_{n*}), \tag{B.3}$$

where $\mathrm{H}_* \coloneqq \lim_{n\to\infty} N^{-1} \bar{\mathrm{H}}(\boldsymbol{\alpha}_*, \beta_{n*})$. $\beta_{n\sharp}$ can be seen as a projection of β_{n*} by concentrating out the fixed effects. When the link function is correctly specified, $(\boldsymbol{\alpha}_*, \beta_{n*}) \equiv (\boldsymbol{\alpha}_0, \beta_0)$, thus $\beta_{n\sharp} \equiv \beta_{n*} \equiv \beta_0$ because $\mathbb{E}[\mathbf{s}_{1,0}] = 0$ and $\mathbb{E}[s_{2,0}] = 0$ imply $\bar{s}_{n,0} \equiv 0$. Furthermore, our OS and BG estimators in the misspecified case share similar asymptotic properties as their counterparts under correct specification, except that they now center on the projected pseudo-true value $\beta_{n\sharp}$ instead of β_0.

For the next theorem, define b_* as b_0 but with $(\boldsymbol{\alpha}_*, \beta_{n*})$ and the misspecified link function $q(\cdot)$. The asymptotic covariance matrix Γ_* is

$$\Gamma_* \coloneqq \lim_{n\to\infty} N^{-1} \mathrm{H}_*^{-1} \begin{bmatrix} \mathrm{I}_{22*} + \mathbf{H}_{12*}^\top \mathbf{H}_{11*}^{-1} \mathbf{I}_{11*} (\mathbf{H}_{11*}^{-1} \mathbf{H}_{12*})^\top \\ - \mathbf{H}_{12*}^\top \mathbf{H}_{11*}^{-1} \mathbf{I}_{12*} - (\mathbf{H}_{12*}^\top \mathbf{H}_{11*}^{-1} \mathbf{I}_{12*})^\top \end{bmatrix} (\mathrm{H}_*^{-1})^\top. \tag{B.4}$$

Theorem B.2 (OS and BG Estimation under Misspecified Link Function)**.** *Suppose all the bounds in Lemma 4 hold for each element of* $\mathbf{H}_{12}^\top \mathbf{H}_{11}^{-1}$. *If Assumptions 1–4 and 7 are satisfied, then*

$$\sqrt{N}(\widehat{\beta}_{\mathrm{OS}} - \beta_{n\sharp}) + \mathrm{H}_*^{-1} b_* \xrightarrow{d} \mathcal{N}(0, \Gamma_*) \quad \text{and} \quad \sqrt{N}(\widehat{\beta}_{\mathrm{BG}} - \beta_{n\sharp}) \xrightarrow{d} \mathcal{N}(0, \Gamma_*).$$

Theorem B.2 shows that $\widehat{\beta}_{\mathrm{BG}}$ is robust: under correct specification it centers on β_0 and attains the CRLB; otherwise it centers on the projected pseudo-true value $\beta_{n\sharp}$ with no asymptotic bias. The covariances Ω_* and Γ_* are consistently estimated by plugging $(\widehat{\boldsymbol{\alpha}}, \widehat{\beta})$ into the formulas of Theorems B.1 and B.2.

B.1 Proofs of Theorems B.1 and B.2

Proof. Both proofs proceed in parallel to their correctly specified counterparts (Theorems 1, 2, and 3), with $(\boldsymbol{\alpha}_0, \beta_0)$ replaced by the pseudo-true parameters $(\boldsymbol{\alpha}_*, \beta_{n*})$ and $(\boldsymbol{\alpha}_*, \beta_{n\sharp})$, respectively. The existence and uniqueness of $\boldsymbol{\alpha}_*$ follows from Lemma 1 applied to the misspecified link function $q(\cdot)$ under Assumption 7.

For Theorem B.1: the consistency argument is identical to that of Theorem 1, with $\bar{S}_n(\beta)$ replaced by $\widetilde{S}_n(\beta)$ from Assumption 7. The asymptotic normality follows the same Taylor expansion, except that the variance of $y_{ij} - q_{ij}(\boldsymbol{\alpha}_*, \beta_{n*})$ under the true DGP is $p_{ij,0}(1 - p_{ij,0})$ (not $q_{ij}(1 - q_{ij})$), yielding the sandwich-form variance Ω_* in place of Ω_0.

For Theorem B.2: the one-step expansion parallels Theorem 2, with the Hessian $\mathbf{H}$ replacing the information matrix $\mathbf{I}$ (since the information equality $\mathbb{E}[\mathbf{H}] = -\mathbf{I}$ fails under misspecification). The population concentrated score $\bar{s}_n(\boldsymbol{\alpha}_*, \beta_{n*}) \neq 0$ shifts the pseudo-true target from β_{n*} to $\beta_{n\sharp}$ as defined in (B.3). The bagging argument is unchanged. □

C Monte Carlo Simulations

This section complements the baseline results in Section 5 of the main text. Section C.1 illustrates the non-concavity challenge faced by direct MLE. Section C.2 presents extended NTU results: fixed-effect recovery, the APE estimator, link-function misspecification, and sparser networks. Section C.3 briefly summarizes analogous TU results.

C.1 Non-Concavity of MLE in the Fixed Effects

The major challenge of directly invoking MLE is the non-concavity of the log-likelihood function in the high-dimensional fixed effects $\boldsymbol{\alpha}$. To illustrate this, Figure C.1 compares the distributions of $\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|_2$ obtained by JMM and MLE, both initialized at the same starting values.[1] Both panels draw $\boldsymbol{\alpha}_0$ from bounded supports but with structures that expose the non-concavity problem. The left panel uses a symmetric bimodal design $\alpha_{i0} \overset{i.i.d.}{\sim} 0.5\,U(-3,-1) + 0.5\,U(1,3)$, with a gap at zero; the right panel considers a hub–periphery design $\alpha_{i0} \overset{i.i.d.}{\sim} 0.9\,U(-1.5,-0.5) + 0.1\,U(3,4)$,

[1] MLE is implemented using the L-BFGS-B algorithm in Python's SciPy package.

in which a small fraction of high-degree hubs coexists with a low-degree periphery. In both settings the contrast is stark: JMM's RMSE distribution is tightly concentrated near zero, whereas the bulk of MLE's distribution shifts to substantially larger values, indicating that MLE often fails to consistently recover $\boldsymbol{\alpha}_0$ from the same starting point. The hub–periphery case is more extreme, with MLE errors reaching nearly 10 in some replications. These results highlight that, when the fixed effects are high-dimensional, the non-concavity of the log-likelihood poses a serious practical obstacle for MLE even under bounded support, which our moment-based approach effectively circumvents.

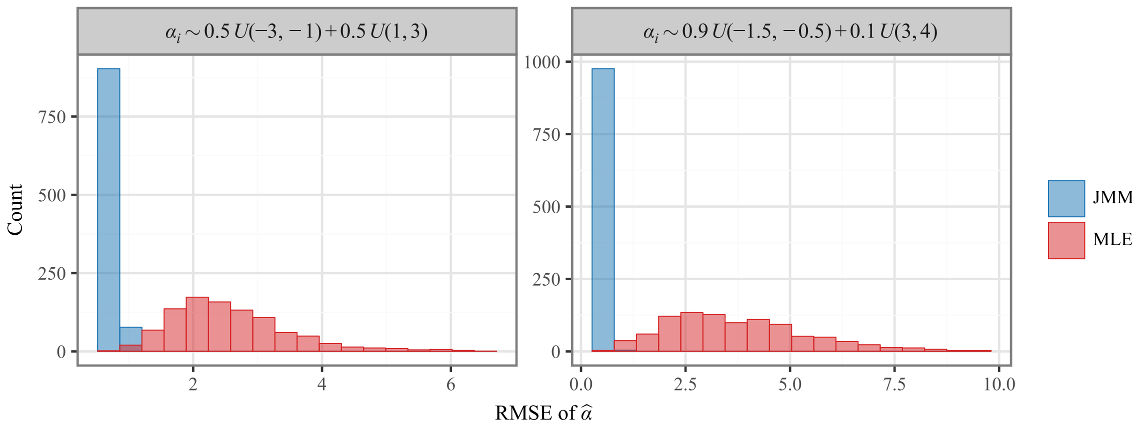

Figure C.1: Distributions of RMSE of $\widehat{\boldsymbol{\alpha}}$ under two NTU DGPs ($n = 100$)

C.2 Extended NTU Results

This section reports further finite-sample results under the NTU specification of Section 5.

Fixed-effect recovery. To examine estimation performance in the high-dimensional setting, where $\boldsymbol{\alpha}_0$ is an $n \times 1$ vector, we additionally consider $n = 500$ and 1,000. Figure C.2 reports histograms of $\widehat{\alpha}_i - \alpha_{i0}$ for $i \in \mathcal{I}_n$. Across all sample sizes, the histograms are centered around zero. As n increases, the estimation accuracy of $\widehat{\alpha}_i$ improves, and the dispersion of the estimation errors contracts toward zero, consistent with the theory.

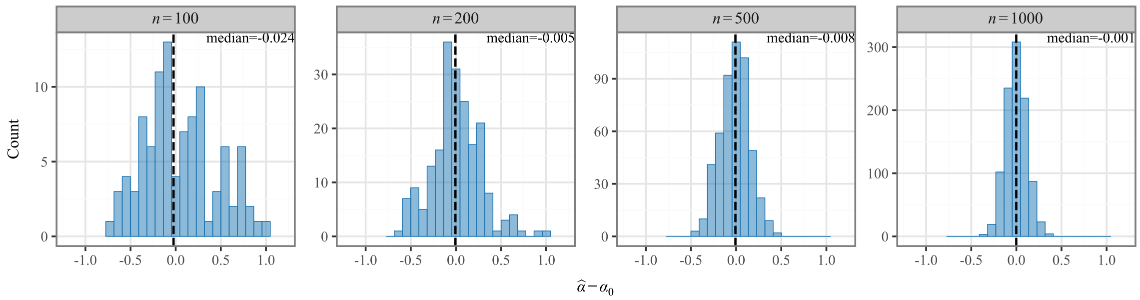

Figure C.2: Histograms of $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0$ under NTU for different n

APEs. Table C.1 summarizes the APEs for each coordinate of X_{ij} defined in (11). Our plug-in estimator performs well with respect to RMSE and coverage probabilities. When applied to the estimation of APEs, the split-network jackknife bagging method does not yield meaningful improvement. As predicted by Theorem 4, the asymptotic bias in estimating APEs is $O(n^{-1})$, which is negligible relative to the APE estimator's $O(n^{-1/2})$ standard error.

Table C.1: NTU estimation results for the APEs

	$n = 100$				$n = 200$			
	Plug-in		Bagging		Plug-in		Bagging	
	$X_{ij,1}$	$X_{ij,2}$	$X_{ij,1}$	$X_{ij,2}$	$X_{ij,1}$	$X_{ij,2}$	$X_{ij,1}$	$X_{ij,2}$
Mean Bias	-0.14	0.39	-0.16	0.14	-0.00	0.19	-0.00	0.09
Median Bias	-0.04	0.46	-0.09	0.21	0.04	0.15	0.06	0.08
Standard Deviation	2.75	2.79	2.89	2.85	1.32	1.42	1.37	1.45
Mean Standard Error	2.78	2.88	2.78	2.88	1.40	1.46	1.40	1.46
Mean Absolute Bias	2.18	2.26	2.28	2.29	1.04	1.15	1.08	1.16
Median Absolute Bias	1.78	1.87	1.91	1.88	0.84	0.97	0.89	0.99
RMSE	2.75	2.82	2.90	2.86	1.32	1.43	1.37	1.45
90% Coverage Rate	90.8	90.2	89.2	89.2	91.8	90.6	90.8	90.8
95% Coverage Rate	95.0	95.0	94.4	95.2	96.2	95.7	95.4	96.1

Note: All values have been multiplied by 100; true values of APEs are calibrated from a simulation with $n = 10,000$ agents.

Link-function misspecification. Table C.2 presents the results for estimating the homophily coefficients under misspecification of the distribution of ϵ_{ij}. We draw ϵ_{ij} from the standard normal distribution, but "mistakenly" specify the logistic link function in the estimation. Following Theorems B.1 and B.2, we compare JMM to β_{n*} and OS/BG to $\beta_{n\sharp}$ defined in (B.3), and find that the results are satisfactory. The performance of our BG estimator dominates other estimators in terms of bias, variance, and coverage probabilities, highlighting the efficacy and importance of proper bias-correction procedures.

Table C.2: NTU estimation results for β_0 under link function misspecification

$n = 100$	JMM		OS		BG	
	β_1	β_2	β_1	β_2	β_1	β_2
Mean Bias	5.92	-5.34	5.36	-5.14	0.15	-0.02
Median Bias	5.99	-4.83	5.37	-4.91	0.24	0.11
Standard Deviation	6.16	15.36	6.16	15.34	5.98	14.93
Mean Standard Error	6.24	15.55	6.20	15.39	6.20	15.39
Mean Absolute Bias	7.03	12.81	6.66	12.76	4.78	11.89
Median Absolute Bias	6.29	10.18	5.91	10.44	4.02	9.84
RMSE	8.55	16.26	8.16	16.17	5.99	14.93
90% Coverage Rate	75.8	88.6	78.0	88.5	91.7	90.8
95% Coverage Rate	84.0	94.2	86.5	94.1	96.7	95.6
$n = 200$	JMM		OS		BG	
	β_1	β_2	β_1	β_2	β_1	β_2
Mean Bias	2.67	-2.69	2.49	-2.54	-0.08	0.01
Median Bias	2.66	-2.62	2.47	-2.32	-0.11	0.24
Standard Deviation	3.06	7.44	3.04	7.43	2.99	7.34
Mean Standard Error	3.05	7.51	3.04	7.44	3.04	7.44
Mean Absolute Bias	3.28	6.28	3.16	6.24	2.38	5.90
Median Absolute Bias	2.89	5.18	2.78	5.29	1.98	5.06
RMSE	4.06	7.91	3.93	7.86	2.99	7.34
90% Coverage Rate	78.5	88.5	79.3	89.0	90.0	91.2
95% Coverage Rate	86.6	94.0	86.7	94.1	94.9	95.0

Note: All values have been multiplied by 100.

Sparser networks. We examine the performance of the method in networks with fewer links on average. To this end, we lower all α_i's by one, resulting in a network density of 8.6%. As reported in Table C.3, network sparsity worsens the performance of all estimators. Nevertheless, the BG estimator continues to outperform the others across all metrics.

Table C.3: NTU estimation results for β_0 under a sparser network

$n = 100$	JMM		OS		BG	
	β_1	β_2	β_1	β_2	β_1	β_2
Mean Bias	4.25	-7.68	4.76	-4.24	-0.46	0.70
Median Bias	4.22	-8.26	4.78	-4.77	-0.35	0.01
Standard Deviation	7.45	18.58	7.45	18.66	7.08	17.91
Mean Standard Error	7.39	17.91	7.36	17.81	7.36	17.81
Mean Absolute Bias	6.90	16.21	7.11	15.45	5.69	14.37
Median Absolute Bias	5.86	14.13	6.17	13.25	4.97	11.99
RMSE	8.58	20.11	8.84	19.14	7.09	17.93
90% Coverage Rate	85.8	85.4	83.5	87.6	91.5	90.4
95% Coverage Rate	91.7	92.2	90.1	93.2	96.5	95.1
$n = 200$	JMM		OS		BG	
	β_1	β_2	β_1	β_2	β_1	β_2
Mean Bias	2.14	-3.91	2.35	-2.45	-0.14	-0.04
Median Bias	1.99	-3.95	2.19	-2.44	-0.30	0.02
Standard Deviation	3.65	8.74	3.65	8.72	3.55	8.53
Mean Standard Error	3.59	8.72	3.58	8.68	3.58	8.68
Mean Absolute Bias	3.37	7.61	3.47	7.18	2.85	6.77
Median Absolute Bias	2.79	6.19	2.95	5.90	2.48	5.68
RMSE	4.23	9.58	4.34	9.06	3.56	8.53
90% Coverage Rate	84.0	86.2	82.3	87.9	91.4	90.9
95% Coverage Rate	90.3	93.0	88.8	94.0	95.2	95.2

Note: All values have been multiplied by 100.

C.3 Transferable-Utility Extended Results

For the TU specification, we also conducted analogous exercises at $n = 100$ and $n = 200$ under both logistic and probit links, covering fixed-effect recovery, the APE estimator, link-function misspecification, and sparser networks. The qualitative conclusions match those reported for NTU: the BG estimator substantially reduces the bias of JMM and OS while preserving coverage close to the nominal level, and the APE results are consistent with the $O(n^{-1})$ bias bound of Theorem 4. The detailed tables are omitted for brevity and are available from the authors upon request.

D Additional Details for Empirical Applications

This section provides detailed data descriptions and summary statistics for the two empirical applications in Section 6 of the main text, together with the distribution of the estimated individual fixed effects $\widehat{\boldsymbol{\alpha}}$ from the Nyakatoke NTU specification (Figure C.3).

D.1 Townsend Thai Village Networks

The Townsend Thai monthly panel is a long-running household survey covering rural and peri-urban villages in four provinces of Thailand. We use the subsample of $V = 16$ villages for which the transaction-level network data have been compiled by Kinnan, Samphantharak, Townsend, and Vera-Cossio (2024). Each village has an average of 44 households (min 24, max 50), and the data record monthly transactions across multiple domains from September 1998 through December 2012. We aggregate monthly transactions to the annual level, yielding a pooled dyad sample of $N = \sum_{v=1}^{16} \binom{n_v}{2} = 15{,}641$ observations in each year.

We construct three binary dyadic outcome variables from the transaction records:

(i) a *financial* link, equal to 1 if the two households exchange gifts or informal loans/repayments during the year;

(ii) an *operations* link, equal to 1 if they transact in production inputs, intermediate goods, or output sales/purchases; and

(iii) a *labor* link, equal to 1 if one household hires labor from the other or they exchange labor.

We treat each resulting binary pairwise indicator as $Y_{ij,v}$ in (1) under the TU specification. Because our theory requires a dense-network regime, for each outcome we select the year with the highest pooled density: year 3 for the financial network, year 2 for the operations network, and year 3 for the labor network.

Our covariates are three dyadic variables constructed from baseline household characteristics, together denoted X_{ij}:

(i) $X_{1,ij} = \ln(\|D_i - D_j\|_2)$, the (ln) demographic distance, where D_i is a vector of baseline household composition and head characteristics (number of males, females, adults, children, total size, mean age, and head's sex, age, and education);

(ii) $X_{2,ij} = |\ln(W_i) - \ln(W_j)|$, the absolute (ln) difference in baseline net worth W_i (aggregated from the household financial accounts); and

(iii) $X_{3,ij} \in \{0, 1\}$, an indicator for kinship based on the reported relationship codes in the household roster.

Summary statistics for the three outcome variables and the covariates, pooled across villages, are reported in Table C.4.

Table C.4: Summary statistics for the Townsend Thai village networks (pooled across $V = 16$ villages)

Variable	Mean	Std. Dev.	Min	Max
Financial link (dependent variable, year 3)	0.0146	0.1201	0	1
Operations link (dependent variable, year 2)	0.0536	0.2252	0	1
Labor link (dependent variable, year 3)	0.0779	0.2681	0	1
(ln) Demographic difference	5.4840	0.6193	1.7132	7.0546
(ln) Net-worth difference	2.0698	4.0895	0	32.7114
Kinship	0.0330	0.1786	0	1

D.2 Nyakatoke Risk-Sharing Network

The network data of Nyakatoke, located in the Kagera Region of Tanzania, covers a small Haya community of all 119 households. We investigate how important wealth difference, distance, and blood or religious ties are in deciding the formation of risk-sharing links among local residents. The dataset includes the following variables: (i) whether or not two households are linked in the insurance network, (ii) total USD assets and religion of each household, (iii) kinship and distance between households. To define the dependent variable *link*, each household was asked:

"*Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labor?*"

The data contains three answers of "bilaterally mentioned", "unilaterally mentioned", and "not mentioned" between each pair of households. Considering the question is about whether one can rely on the other for help, we interpret both "bilaterally mentioned" and "unilaterally mentioned" as indicating that they are connected in this undirected network. This coding is broader than literal bilateral consent and should be interpreted as a proxy for an underlying risk-sharing relationship. In the context of village economies, these links are unlikely to be formed through explicit side-payment transfers, making NTU the more natural benchmark.

We estimate the coefficients for three regressors: *wealth difference*, *distance* and *tie* between households. *Wealth* is defined as the total assets in USD owned by each household, including livestock, durables and land. *Distance* measures how far away two households are located in meters. *Tie* is a discrete variable, with the value "3" if members of one household are parents, children and/or siblings of members of the other household, "2" if nephews, nieces, aunts, cousins, grandparents and grandchildren, "1" if any other blood relation applies or if two households share the same religion, and "0" if no blood or religious tie exists. Following the literature, we take the natural logarithm on *wealth* and *distance*, and we construct the *wealth difference* variable as the absolute difference in *wealth*, i.e.,

$$X_{ij} = \left(\left| \ln\left(\text{wealth}_i\right) - \ln\left(\text{wealth}_j\right) \right|, \ \ln\left(\text{distance}_{ij}\right), \ \text{tie}_{ij} \right)^\top .$$

Five households in the data have no information on *wealth* and/or *distance*. We drop these observations, resulting in a sample of $n = 114$ households and $N = 6{,}441$ dyadic observations. Table C.5 reports the summary statistics.

Table C.5: Summary statistics for the Nyakatoke network

Variables	Mean	Std. Dev.	Min	Max
link	0.0733	0.2606	0.0000	1.0000
(ln) wealth difference	1.0365	0.8227	0.0004	5.8898
(ln) distance	6.0553	0.7092	2.6672	7.4603
tie	0.4260	0.6123	0.0000	3.0000

Figure C.3 plots the distribution of the estimated individual fixed effects $\widehat{\boldsymbol{\alpha}}_i$ from the NTU specification reported in Section 6.2.1 of the main text. Most estimated fixed effects lie in the range $[2, 5]$, although some exceed this range, reflecting unobserved household heterogeneity.

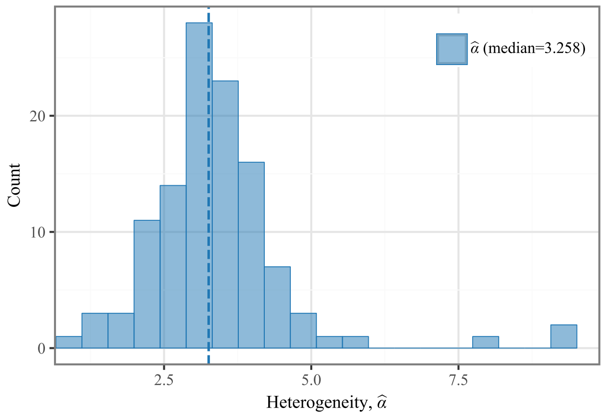

Figure C.3: Histogram of $\widehat{\boldsymbol{\alpha}}$ for the Nyakatoke network

References for Supplemental Material

Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Chatterjee, S., P. Diaconis, and A. Sly (2011). Random graphs with given degree sequence. *Annals of Applied Probability 21*(4), 1400–1435.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Kinnan, C., K. Samphantharak, R. Townsend, and D. Vera-Cossio (2024). Propagation and insurance in village networks. *American Economic Review 114*(1), 252–284.

van der Vaart, A. W. (2000). *Asymptotic Statistics*, Volume 3. Cambridge University Press.

Yan, T. (2019). Approximating the inverse of a diagonally dominant matrix with positive elements. arXiv Preprint arXiv:1902.00668.

Yan, T., B. Jiang, S. E. Fienberg, and C. Leng (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association 114*(526), 857–868.

Yan, T. and J. Xu (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika 100*(2), 519–524.